PHD

Structural studies on glycosyltransferases

Jamaluddin, Haryati

Award date:
2007

Awarding institution:
University of Bath

Structural Studies on Glycosyltransferases

Volume 1 of 3

Haryati Jamaluddin

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

October 2007

Copyright

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis is dedicated in loving memory of my father, who had always told me that hard work is never easy but the end product is always worth while. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

LIST OF ABBREVIATIONS

α3GT	α-1,3 galactosyltransferase
Å	Ångström
ADA	N-(2-Acetamido)iminodiacetic Acid
B-factor	temperature factor
CΔ3	Mutant with last 3 residues of the C-terminus truncated
CAZy	Carbohydrate Active Enzymes
CCP4	Collaborative Computational Project Number 4
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
FG	Forssman Glycolipid
FS	Forssman Synthase enzyme
Gal/gal	Galactose
Glc	Glucose
GOL	Glycerol
GT	Glycosyltransferase
GTA	Human N-acetylgalactosaminyltransferase (Human ABO blood group enzyme)
GTB	Human Galactosyltransferase (Human ABO blood group enzyme)
HCl	Hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
iGB3-S	Isoglobotrihexosylceramide Synthase
M	Molarity
MES	2-(n-morpholino)-ethanesulfonic acid
Mn	Manganese (MN)
$MnCl_2$	Manganese chloride
MPD	2,4-dimethyl pentane diol
NAG	*N*-acetylgalactosamine (GalNAc)
NAL	*N*-acetyllactosamine (LacNAc)

NaOH	Sodium hydroxide
PDB	Protein Data Bank
PEG	Polyethylene glycol
pNP	P-nitrophenyl
r.m.s.d	Root mean square deviation
SRS	Synchrotron Radiation Source
TRS	Trishydroxymethylaminomethane (Tris)
UDP	Uridine 5'-diphosphate
UDP-2F-gal	UDP 2'-fluorogalactose (UP1)
UDP-gal	UDP-galactose (GDU)
UDP-glc	UDP-glucose

List of Amino acids		
Amino acids	**3-letter code**	**1-letter code**
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Materials and Methods

All chemicals mentioned in this thesis were obtained from Sigma unless otherwise stated.

All figures were made using Pymol (DeLano, W. L. (2003) The PyMOL molecular graphics system. DeLano Scientific, San Carlos, CA, http://www.pymol.org)

Abstract

Specific hetero-oligosaccharides on glycoproteins and glycolipids play important roles in cell-cell and cell-matrix interactions, affect the stability and structure of proteins, and are epitopes for immune system recognition. Glycosyltransferases (GTs) are enzymes that catalyse the synthesis of these oligosaccharides. These enzymes are abundant in nature and represent one of the most widespread in the living world. They can be found in organisms as different as bacteria, yeast, plants and most of the living animals. Structural information on glycosyltransferases would facilitate in understanding and addressing the structural basis for the catalysis of specific glycoconjugates in the cell. Furthermore, structural data would be useful for the design of inhibitors and to assist the engineering of new catalysts for the enzymatic synthesis of natural and novel glycoconjugates for therapeutic use.

The bulk of this study is concentrated on structural studies of bovine α-1,3-Galactosyltransferase (α3GT) which is an enzyme that catalyzes the transfer of Galactose from UDP-α-D-galactose into an α-1,3 linkage with β-galactosyl groups in glycoconjugates. α3GT is a member of a family of metal-dependent retaining glycosyltransferases including the histo-blood group A and B synthases. The enzyme is expressed in many mammalian species but is absent from humans, apes and old world monkeys. This presents a barrier against xenotransplantation which is one of the major approaches to dealing with the limited supply of donor organs. Hence understanding the catalytic mechanism of this enzyme would have significant biological implications.

Chapter 1 of this thesis outlines the general principles of X-ray crystallography and gives an overview of the practice of this technique. Chapter 2 provides a general biological and structural background to glycosyltransferases, the metal retaining family 6 glycosyltransferases and then move on to specifically describe α-1,3-Galactosyltransferase. Chapter 3 presents the majority of the work which is on structural studies of α-1,3-Galactosyltransferase. This chapter is divided into four sub chapters starting with the description of protein preparations and crystallisation of the wild type and mutant proteins. The second sub chapter discusses the conformational dynamics of the enzyme by structural studies of 'apo' forms of the enzyme and C-terminal mutants of

the enzyme. Subchapter three describes the structural studies carried out on active site mutants of α3GT in order to better understand the catalytic mechanism of the enzyme. Subchapter 4 discusses binding studies of wild type α3GT with an alternative acceptor substrate *p*-Nitrophenyl galactoside (pNPGal) and also discusses modelling studies of the enzyme with the inhibitor, U66. Chapter four of the thesis describes attempts to crystallise canine Forssman synthase which is an enzyme that also belongs to the metal retaining family 6 glycosyltransferases. This chapter presents the crystallisation trials which had been carried out in order to obtain a diffraction quality crystal. A brief final section comments on the status of, and future for, research in these areas.

Contents

Acknowledgements

It is a pleasure to thank the many people who made this thesis possible. Firstly I would like to extend my deepest gratitude to my Ph.D supervisor, Professor Ravi Acharya. With his enthusiasm, his inspiration, patience and determination in helping me succeed in this research field. A big thank you also goes to Miss Percy Tumbale and Professor Keith Brew from Florida Atlantic University, USA for the endless supply of highly pure proteins without which this thesis would not have been possible. I am also thankful to the staff at the SRS Daresbury and EMBL Hamburg for their help, assistance and beam time at the synchrotrons.

I would like to thank the kind, knowledgeable and friendly people that helped me daily in the lab/office. My office mates, Dr Umesh Singh, Mr Amit Sundriyal and Dr. Talat Jabeen for long talks, endless supply of 'biskuts' and for just being there for me when I needed an opinion or advice. Thank you to Dr. Shalini Iyer, Dr Hazel Corradi and Dr Matthew Baker for all the tips and help during my stressful manuscript and thesis writing period. Thank you to my other office mates, including Dr Daniel Holloway, Dr Kenneth Holbourn, Miss Eirini Mitsiki, Miss Nadia Kazakou, Dr. Nethaji Thiyagarajan, and Mr Peter Wilson who were each a great help in their own way.

In addition to the people in my office, I have been lucky enough to have the support of many good friends. Ph.D life would have been more stressful without my close friends and fellow Ph.D students, Mr Kui Soon Lee, Miss Ajaraporn Sriboonlert and Miss Lihua Zhang. Thank you for the many lunches together and for the wonderful company. Special thank you goes to my 'ex-housemate' and close friend, Miss Anna Tai for the endless moral support and entertaining talk. Finally, I would like to thank those closest to me, whose presence helped make the completion of my graduate work possible. Thank you to Sharifah Nurul Ain Syed Nasir and Syed Firdzuan for accepting me with open arms when I needed a place to stay. My gratitude extends to everyone else who made me feel loved and at home here in Bath, to Siti Nurasyikin Hamzah, Badiatul Adawiyah Zubir, Wan Khaizuri Wan Seman, Amin Sahat and Kak Zakiah Ahmad, just to name a few of these lovely people. Special thanks to my adopted family;

Abang Najib Masshor, Kak Zalima Drus, Nayim Masshor and Nashua Masshor for being extremely kind to me especially in the last month of my stay in Bath.. Most of all, I would like to thank my family and especially my mother for her love and absolute confidence in me.

Lastly I would like to thank Universiti Teknologi Malaysia, The Malaysian Government, The British Federation of Women Graduates and University of Bath for the financial support which made it possible for me to pursue my Ph.D studies.

CHAPTER 1

Protein Crystallography

1. X-ray Crystallography

1.1 General Background

Crystallography provides the most direct way of visualizing the images of molecules. Through the method of X-ray crystallography thousands of 3-dimensional structures of proteins and nucleic acids have been elucidated. These give detailed information about their activity, their mechanism for recognizing and binding substrates and effectors, and the conformational changes they might undergo (Blow, 2002). Armed with this information we can design novel drugs that target a particular protein, or rationally engineer an enzyme for a specific industrial process. Over the years, with the advent of increasingly faster computers, crystallography software and improved X-ray technology has lead to many improvements in the speed and accuracy of structure solution (Helliwell, 1992). In turn, the information from many completed molecular structures has allowed us to further improve our interpretation of X-ray diffraction data. There are several steps towards the elucidation of a given protein structure through X-ray crystallography and they will be discussed in this chapter. The various steps towards protein structure determination are:

- Protein preparation
- Crystallisation
- Data collection and processing
- Phase determination
- Electron density map calculation
- Model building and refinement
- Structure validation and analysis

1.2 Protein preparation

The first step to obtaining a crystal structure of a given protein is acquiring pure protein for crystallisation. The availability of pure protein is usually detrimental towards obtaining diffraction quality crystals hence acquiring high quality

diffraction for structure elucidation. The aim of protein purification ideally is to produce a pure (>99%), homogenous, correctly folded, monodisperse, active and stable protein sample. The protein sample should also be readily produced in large quantities for the means of crystallisation trials. There are several aspects which should be considered when purifying proteins for crystallography and they include the source of the protein (either wild type or recombinant), the expression system, fusion proteins and cleavage of tag, purification method and assessment of purity. Purity of a protein sample is essential in obtaining high quality diffraction data as low purity can be a barrier to crystal growth. Contaminating species can interfere with the bonding interactions needed to propagate the crystal lattice hence leading to a reduction in crystal formation. However, the protein samples available are not usually ideal for crystallisation as they could be difficult to purify and may only be produced in low concentration or have low stability. In order to overcome this, protein can be stored in buffers containing high concentration of glycerol (50%) and then it can be subjected to buffer exchange and concentrated to the appropriate concentration (5-20 mg/ml) prior to crystallisation.

1.3 Protein crystallisation

Crystals are regular, three-dimensional arrays of atoms, ions, molecules or molecular assemblies. The aim of a crystallisation experiment is to produce diffraction quality crystals which are the basis of X-ray structure determination. It is also sometimes the primary difficulty in the determination of a protein structure. The process involves controlled precipitation from aqueous solution under conditions that do not denature the protein (Rhodes, 2000). Proteins are structurally dynamic systems, often micro-heterogenous, whose properties are influenced by environmental conditions such as pH, temperature, ionic strength and a number of other factors. Protein purity and homogeneity is essential to the growth of single protein crystals. Crystallisation of macromolecules is a process involving three major stages: nucleation, growth and cessation of growth. It is of utmost importance for crystallisation to bring the protein

to a supersaturated state (Figure 1.1), which will force the molecules to aggregate together and ultimately form the solid state which is the crystal.

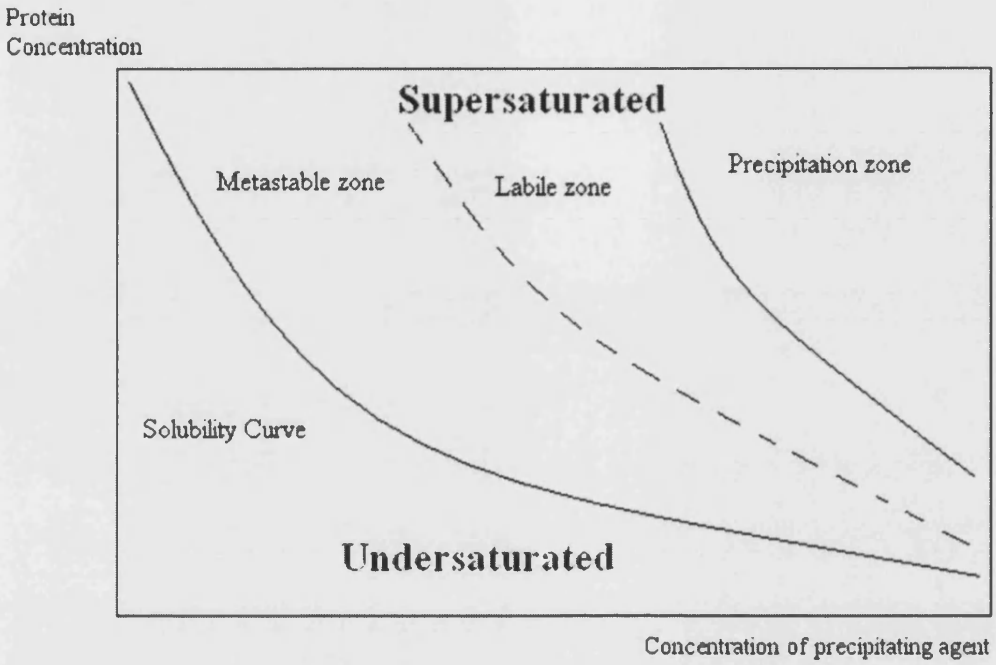


Figure 1.1: Solubility diagram. It represents the different steps occurring during crystal formation. The supersolubility curve separates the labile region, where nucleation occurs, from the metastable region, where crystals grow.

The precipitants normally used for crystallisation are salts, polyethylene glycols or organic solvents like ethanol, isopropanol or methylpentane diol. At low ionic strength (low ionic concentration), the solubility of protein is higher if the amount of electrolytes is increased –termed "salting in". At high ionic strength the ions start to compete with each other for water molecules, resulting in a decrease in solubility. This process is known as "salting out". The crystallographer can shift the equilibrium to supersaturation by increasing or reducing the ionic strength of the protein solution. The most widely used precipitant is polyethylene glycol (PEG) because it is a powerful precipitant and a weak denaturant. PEG is often available in a variety of average molecular masses such as PEG 4000 which has an average molecular mass of 4000 daltons. As shown in Figure 1.1, crystal growth can be divided into two steps. First, a spontaneous nucleus formation occurs in the

supersaturation area followed by formation of small aggregates. After the critical amount of aggregated molecules is surpassed, the crystal growth becomes an energetically privileged process. Crystal growth always needs a lower degree of supersaturation than nucleus formation. Crystals should grow slowly to achieve the maximum possible internal order. It is clear that crystal morphology is not a direct indication of crystal quality. ''Good looking'' crystals can have disordered crystal packing limiting their diffraction properties. Contrary to this, some ''morphologically poor'' crystals can give a positive surprise when they are measured. Different methods for growing crystals, as batch crystallisation, dialysis, liquid-liquid diffusion and vapour diffusion, have been established and developed in the direction of using less amounts of material. The technique most widely used is hanging drop method (Figure 1.2). The general role is that the sample must be pure and homogenous, it should be clean from small undesirable molecules and should contain a population with the same protein conformation. Microheterogenity of the sample can occur due to different reasons: variation in primary structure (genetic mutations), secondary structure (unfolding), tertiary structure (conformers), quaternary structure (oligomerization), partial oxidation of some groups, fragmentation by proteolysis or molecular dynamics of flexible parts(Bergfors, 1999).

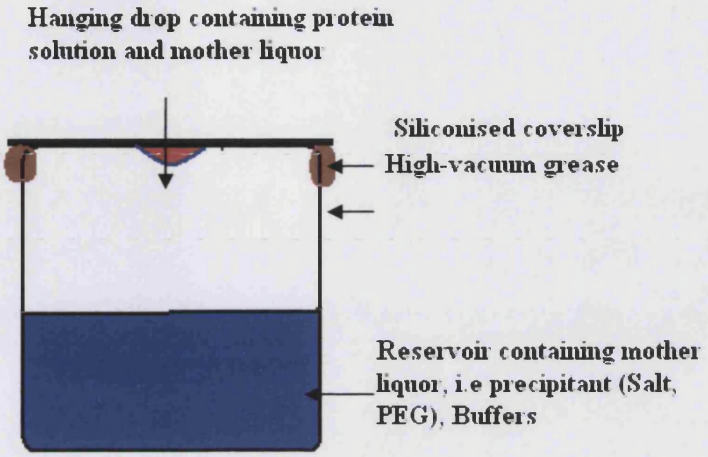


Figure 1.2: A few micro-litre of protein solution is mixed with an equal amount (normally a ratio of 1:1) of reservoir solution, which contains the precipitants and placed as a drop on the coverslip. A thin layer of grease seals the cover slip on the well containing the mother liquor. The most commonly used plates are 24 well linbro plates.

More often than not, waiting for a protein to crystallise can be a long and daunting process. Sometimes spontaneous nucleation needs a helping hand and this is where a technique known as seeding comes into practice. It is profitable to stimulate nucleation by seeding with crystals of the same protein grown under other conditions. Energetically, it is more profitable to add to an already existing crystal plane than it is to create a new nucleus, so seeding increases the probability of nucleation of a new crystal form (Bergfors, 1999). There are three types of seeding technique commonly used: microseeding which is the addition of a very small piece of crushed protein crystal to the crystallisation drop, macroseeding which is the addition of an intact, already grown crystal to the crystallisation drop and streak seeding which is similar to microseeding but involves the use of a fine fibre (often a cat's or rabbit's whiskers are best) to pick up small protein crystal fragment and transfer it into a crystallisation drop. Microseeding is the most reliable form of seeding in which a few small crystals, or one large crystal is broken up to a fine

powder using a needle. A small amount of mother liquor containing this fine powder is added to an aliquot of well solution and then serially diluted. A small quantity of the microseed solution is then added to a pre-equilibrated protein crystallisation drop. For streak seeding, an existing crystal is lightly touched with the fibre to dislodge a few seed and then introduce the seeds into a fresh pre-equilibrated drop by stroking the fibre in a straight line through the drop.

For data collection only best quality crystals should be used in order to obtain structural data. When different crystal forms are obtained, the best diffracting crystals with the highest symmetry should be chosen for further work.

1.4 Principles of X-ray crystallography

1.4.1 X-ray Diffraction of Crystals

X-rays are electromagnetic radiation which are diffracted by even the smallest molecules and X-ray crystallography is unique in its ability to provide a detailed, high resolution picture of molecular structures. In essence, X-rays are used to produce magnified images of protein molecules and the need for X-rays is due to the details of molecular structure being smaller than the wavelength of visible light which is between 400-700nm. If an atomic scale image is to be made, a 'light' of appropriate wavelength must be used. As atoms that are covalently bonded to each other have a bond length of 1-2 Å apart, whilst hydrogen bonds are between 2.5-3.5 Å apart, it is therefore necessary to use wavelengths no larger than a couple of Ångström (1 Ångström= 10^{-8} cm) units (Blow, 2002). Electromagnetic waves of these wavelengths are known as X-rays.

Figure 1.3 depicts the collection of X-ray diffraction data. A crystal is mounted between an x-ray source and an X-ray detector. The crystal diffracts the source beam into many discrete beams which results in a diffraction pattern comprising many reflections. The greater the intensity of the X-ray beam that reaches a particular position, the darker the reflection (Rhodes, 2000). These reflections are measured using a detector such as CCD (charge coupled device), image plate or the area detector. The position and intensity of each reflection is

measured precisely by the detector and is transmitted in digital form to the computer for analysis.

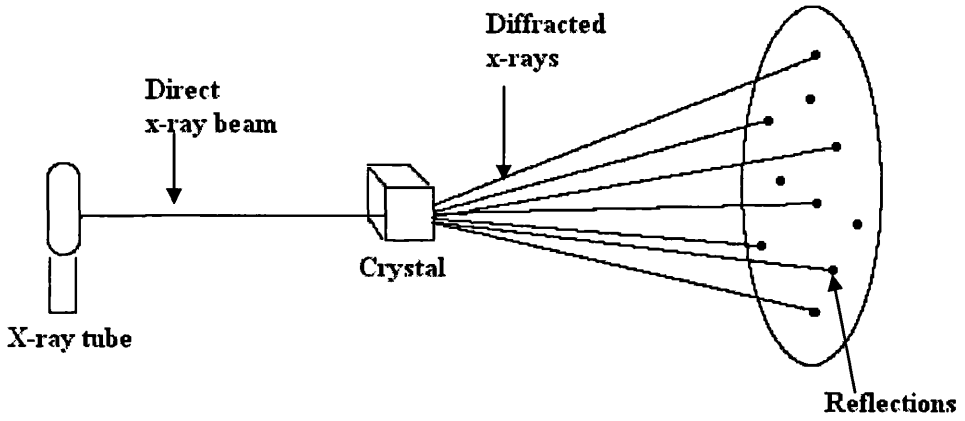


Figure 1.3: Schematic representation of a data collection set up

The position of a reflection can be used to obtain the direction at which a particular beam is diffracted by the crystal while the intensity of the reflection indicates the strength of the diffracted beam that produces a particular spot. The position of each reflection can be given a notation (*hkl*), where *h, k* and *l* are integers in which each component indicates into how many parts this set of planes splits each axis of the unit cell. The central beam position is taken as the origin and assigned the co-ordinates (*h, k, l*) = (0, 0, 0) or *hkl*=000 with the rest of the reflections counted from this origin.

Even though there are many sets of planes throughout the crystal, coherent diffraction can only occur from sets of planes which have spacings that conform to Bragg's Law (Figure 1.4). The Bragg's Law states that for any angle of incident radiation which hits the crystal, coherent diffraction can only exist from the sets of

planes which have spacings that allow a difference in path length between the waves to equal to an integral number of wavelengths (Blundell and Johnson, 1976). This keeps the resulting diffraction in phase hence producing a coherent signal:

$$2d \sin\theta = n\lambda$$

where d is the interplanar distance , θ is the angle of the incident beam while n is the number of wavelength and λ is the wavelength.

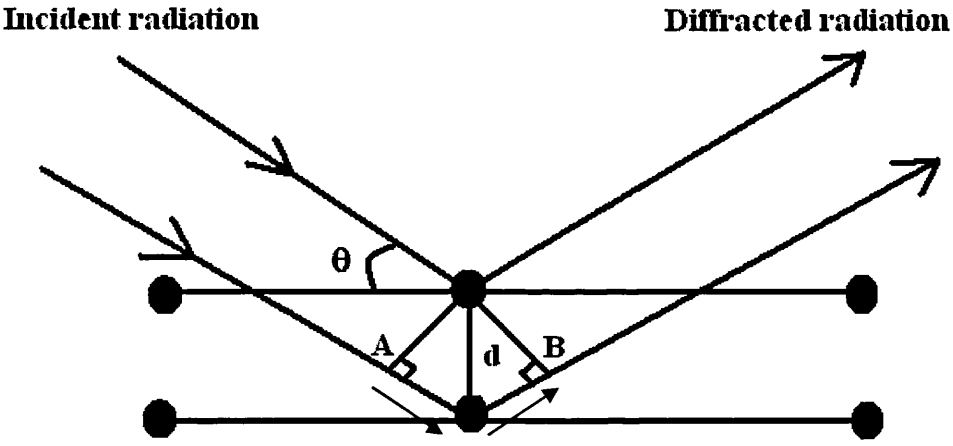


Figure 1.4: Scheme that explains Bragg's law. Two waves that are reflected by two adjacent lattice planes with distance d have a difference in path length that is equal to 2dsinθ, as it can easily be derived from the scheme. A prerequisite for constructive interference is, that this difference in path is an integer multiple of the wavelength used: 2dsinθ=nλ.

The diffraction pattern for any given crystal is a coherent pattern of spots whose properties are related to the dimensions of the crystal. The waves which are diffracted from the real lattice of the crystal forms spots or reflections in the reciprocal lattice of the diffraction pattern. The inverse nature between the real

lattice and the reciprocal lattice can be comprehended by observing Bragg's Law. According to Bragg's Law, the sets of planes which are closer together will only give coherent diffraction at larger angles of incidence. Leading to more dispersed spots on the diffraction pattern whereas the set of planes which are further apart will only give coherent diffraction at narrower angles of incidence resulting in spots closer to the center of the diffraction pattern. Therefore, information for the overall structure of a unit cell contents is concentrated at the center of the diffraction pattern which consists of the lower resolution reflections as these reflections originated from the wider spaced planes in the crystal. The finer details of the unit cell contents will be concentrated in the higher resolution reflections further out in the diffraction pattern as these reflections originated from the more narrowly spaced planes in the crystal. Further description on the concept of resolution is discussed in section 1.3.2. In order to obtain all the information about a given crystal, diffraction patterns from many angles of incident radiation must be recorded. This is achieved by rotating the crystal and this idea can be described best by a simple geometrical construction proposed by Paul Peter Ewald in 1921. Ewald sphere (Figure 1.5) shows that in reciprocal space, there is a sphere of possible reflections which occurs when the crystal and therefore the reciprocal lattice, rotates to bring them into view (Blundell and Johnson, 1976). The radius of this sphere is the inverse of the real space wavelength hence, the shorter the wavelength of the incident radiation, the larger the sphere of reflection will be thus allowing more of the potential reflections to be measured.

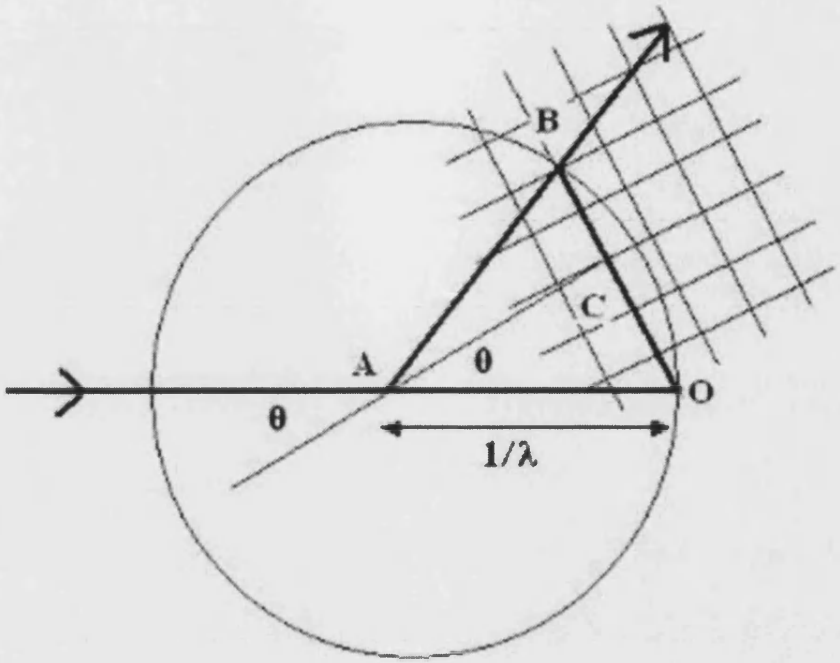


Figure 1.5: The Ewald sphere is a sphere representation of the set of planes which obey the Bragg's Law. A is the crystal center and O represents the point of exit for the non-diffracted beam, the origin of reciprocal space. Radius of the sphere is $1/\lambda$ and when the reciprocal lattice is rotated about O, point B in the reciprocal lattice intersects with the circle resulting in a reflection. This is due to the distance between the origin of the reciprocal space, O, and the diffracted beam being equivalent to $1/d_{hkl}$ in which d is the lattice spacing, hence OC is $1/2d_{hkl}$ which is equivalent to $1/\lambda \sin\theta$. Therefore, $1/2d_{hkl} = 1/\lambda \sin\theta$ which is equivalent to $\sin\theta\ 2d_{hkl} = \lambda$, is Bragg's Law if n = 1. Adapted from (Rhodes, 2000)

1.4.2 The concept of resolution

Resolution of a crystal structure refers to the minimum inter-planar spacing of the reflections included in structure determination. The unit for resolution is represented in Ångström (Å= 10^{-8} cm). Resolution defines the quality of the structure of a macromolecule. The main features of a macromolecule that can be viewed at different resolutions are as follows (Acharya *et al.*, 1995; Blundell and Johnson, 1976):

Resolution	Features
6.0 Å	Outline of the molecule, secondary structure features such as helices and strands, can be identified
3.0 Å	Course of the polypeptide chain can be traced and topology of the folding can be established. With the aid of amino acid sequence, it is possible to place the side chains within the electron density map.
2.0 Å	Main chain conformations can be established with great accuracy. Details of side chain conformations can be interpreted easily without amino acid sequence data. Bound water molecules, metal ions and cofactors can also be identified.
1.5 Å	Individual atoms are almost resolved. It is possible to make out the solvent structure.
1.0 Å	Hydrogen atoms may be visible. It is also possible to see multiple conformations for some side chains.

1.4.3 The unit cell and space group

Although individual atoms diffract X-rays, it is not possible to produce a focussed image of a molecule hence, unfocussed information needs to be collected and converted back to an image using a computational process. A further complication is that single molecules do not diffract X-rays strongly hence, to overcome this obstacle, crystals containing repeating arrays of the molecule are grown to amplify the signal. Crystals comprise a lattice of regularly repeated units known as the unit cells. The unit cell is defined as the minimal structural part that repeats in all three dimensions to build up the crystal. A unit cell is defined by the length of its three axes, a, b, c and the angles between them α, β and γ (Figure 1.6). There are 7 crystal systems for describing unit cells based on their shape in which each of the unit cells can be built up in one of the 14 Bravais lattices to produce the crystal. Within the unit cell, a crystal can contain further symmetry elements, dividing it into several asymmetric units, which form the most basic structural element, which is related to all other identical objects in the unit cell by symmetry elements. The unit cell of a

crystal can contain both rotational symmetry and screw axes, where the asymmetric unit is not only rotated around the axis, but also translated a fraction of the unit cell length. Screw axes are denoted as subscript numbers related to the fraction translation of the unit cell, for example, 2_1 is a two-fold rotation axis with a translation corresponding to half the unit cell length. The geometry of the unit cell together with the possible symmetry operations defines the space group of the crystal. All these elements combined results in only 65 space groups that are feasible for chiral molecules such as proteins. Further details on symmetry and equivalent positions for these space groups, as well as other possible space groups for achiral objects can be found in Volume A of the International Tables for Crystallography (Henry and Lonsdale, 1952).

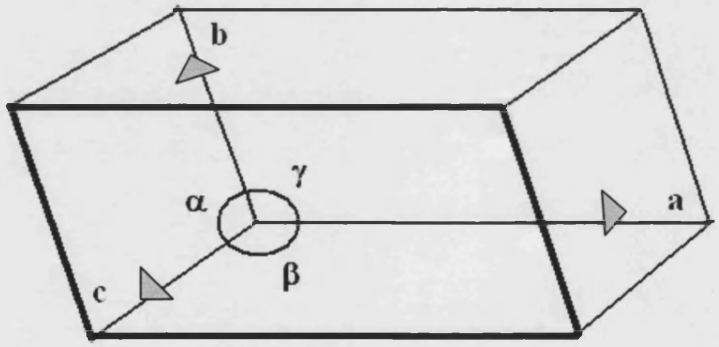


Figure 1.6: The lengths of the unit cell edges are designated as a, b and c and the angle between the axes b and c is α, a and c is β, a and b is γ.

Crystal system	Bravais lattices	Unit cell length (Å)	Angles (°)	Point groups
Triclinic	***P***	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$	1
Monoclinic	***P C***	$a \neq b \neq c$	$\alpha = \gamma = 90$; $\beta > 90$	2
Hexagonal	***P***	$a = b \neq c$	$\alpha = \beta = 90$; $\gamma = 120$	6, 622
Trigonal	***P R***	$a = b = c$	$\alpha = \beta = \gamma \neq 90$	3, 322
Cubic	***P I F***	$a = b = c$	$\alpha = \beta = \gamma = 90$	23, 432
Tetragonal	***P I***	$a = b \neq c$	$\alpha = \beta = \gamma = 90$	4, 422
orthorhombic	***P I C F***	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90$	222

Table 1.2: Seven crystal systems of primitive lattices.

1.5 Diffraction data collection

1.5.1 X-ray source

To perform an X-ray diffraction experiment, we need an X-ray source. The X-rays used for protein crystallography usually come from one of two sources. In crystallography laboratories, a rotating copper anode generator producing an X-ray beam of a characteristic wavelength is often used as a 'home' source. Electrons from a cathode are accelerated towards rotating, a water-cooled, copper anode, resulting in an emission of Cu-Kα radiation at a fixed wavelength of 1.54Å. The X-ray produced is then filtered through a monochromator, made from either nickel or a single crystal. The monochromator functions to effectively remove absorbed radiation and have no contribution to the diffraction pattern. Curved mirrors are then used to deflect the focused X-rays through a low angle which increases the brilliance of the beam (Rhodes, 2000). However, home source radiation is often more suited for characterising crystals and collecting data for crystals that diffract well.

The nature of most protein crystals make them better suited for synchrotron radiation. There are many advantages of using synchrotron sources such as shorter wavelength radiation, tuneable wavelength radiation, greater brilliance and greater flux (Helliwell, 1997). Synchrotron radiation is produced when electrons are accelerated near the speed of light by a magnetic field. During this acceleration, the electron's path is forcibly curved around a large storage ring, in effect causing loss of energy thus resulting in the emission of electromagnetic waves. This electromagnetic radiation, emitted at a tangent to the ring, is then guided and focussed for use in experiments. There is a complete range of electromagnetic radiation produced hence allowing the experimental stations to be designed to offer a variety of wavelengths for diffraction experiments. Shorter wavelengths can be chosen for high resolution experiments and also the ability to choose wavelengths close to the absorption edge of heavy metal for anomalous scattering experiments.

1.5.2 Data collection

Ideally during data collection, we would like to collect the highest possible resolution data set for the crystal. Several factors need to be considered in order to achieve this goal. The diffraction limit depends on the order of the crystal and the wavelength of the X-ray radiation being used. The increasingly common use of synchrotron radiation at wavelengths beyond 1Å has lead to more high resolution structures. The intensity of synchrotron radiation source allows a better signal to noise ratio. However, this same high intensity radiation is also the cause of radiation damage to the crystal. Radiation damage can be reduced by means of collecting data at cryogenic temperatures in the range of 80-100K. In this method, the crystal is picked directly from the drop with a loop of fine nylon or glass fibre and then frozen under a stream of cryogen such as nitrogen. However, not all crystals can be frozen without disruption of the crystal lattice. The crystals would usually need to be soaked in cryoprotectant which is often composed of both the crystallisation solution and a suitable cryoprotectant such as glycerol, polyethylene glycol or MPD (2,4-dimethyl pentane diol). The choice of cryoprotectant is of great importance and varies depending on the crystal. An ideal cryoprotectant should surround the crystal

to prevent the formation of ice crystals that interferes with the diffraction pattern of the crystal. Data collection at cryogenic temperature can be considered as a relatively new method and has become an important tool in crystallography. Previously, data collection was carried out at room temperature in which the crystal is mounted in thin walled glass capillaries in a small amount of the crystallisation solution in order to stop the crystal from drying out. The capillary inner surface is carefully but quickly dried, and a few drops of liquid are placed at each end of the capillary which is then sealed with a high vacuum wax (Blundell and Johnson, 1976). Some of the benefits of cryo-data collection can be summarized as follows:

(i) Reducing primary and secondary radiation damage and increasing the lifetime of the crystal, hence, enabling the collection of an entire dataset from a single crystal.

(ii) Manipulation of smaller, more fragile crystals is permitted due to a reduction of mechanical stress on the crystal.

(iii) The low temperatures minimize thermal motion and allow measurements at higher resolution. Furthermore the cryo method has brought about the establishment of new experimental technique for example; catalytic mechanisms can be deduced by solving the structure of enzymes in complex with reaction intermediates. The low temperature of data collection slows down the reaction mechanism enough to enable a sequence of events or 'snapshots' of catalysis to be observed (Garman, 1997; Rodgers, 1994).

Nevertheless, cryo-data collection is not ideal. During freezing unfavourable changes in the nature of the crystal can be observed. The freezing sometimes alters the alignment of each unit cell with its neighbour resulting in an increased amount of disorder which is reflected in increased mosaicity in the reflection pattern. Unit cell volume also changes at lower temperature causing non-isomorphism between crystals and this can be a problem when collecting native and derivative data sets. The cryoprotectant may also interact with the protein causing conformational changes and may also interfere with ligand binding studies.

As mentioned before, increasing the lifetime of a crystal during data collection is a primary concern for many protein crystallographers and the effort to minimize radiation damage has led to the improvement of detector systems. The logic is that the quicker the data can be recorded for each orientation of the crystal, the less time the crystal will be exposed in the beam, hence, reducing radiation damage. Detectors most commonly used at synchrotrons are the charge-coupled device or CCD detector. This detector accumulates charge on the pixels of the detector where the radiation hits. These charges are systematically passed through fibers to an amplifier before reading. The CCD detector has the advantage of accurate recording of the position and magnitude of the reflections, which are continuously read making data collection fast and accurate.

Once the data collection method has been determined, the crystal must be mounted and aligned in the X-ray beam. The crystal is mounted on a goniometer which is a device that allows the orientation of the crystal to be set precisely. Alignment is usually carried out by rotating the crystal perpendicular to the beam and checking whether the crystal is centered in the beam path at 0^{o}, 90^{o}, 180^{o} and 270^{o} of rotation. Initial orientation of the crystal is usually determined physically. Well-formed crystals usually exhibit distinct faces that are parallel to the unit cell edges and the initial diffraction pattern should be obtained by placing a crystal face perpendicular to the beam. Preliminary examination of the diffraction pattern allows determination of the unit cell dimensions and symmetry. The advancement of cryo data collection and fast detectors along with the automation of the data processing and scaling allows highly redundant, mostly complete datasets to be collected with ease from reasonably good crystals. Most data processing packages such as HKL2000 (Otwinowski and Minor, 1997) and MOSFLM (Leslie, 2001) allow simultaneous data processing with data collection. This provides an advantage as it allows immediate assessment of the quality of the data, allowing changes to be made to the data collection strategy if necessary. Completeness of datasets for very mosaic crystals can be monitored and rapid deterioration of a crystal can be identified immediately allowing the crystal to be changed. In the case of high mosaicity, high redundancy of the data is needed in order to be able to scale a complete data set.

Even though the data collection process can be monitored, several factors should be taken into account before starting a data collection run. From the initial diffraction images, we can predict the exposure time by looking at the strength of the diffraction, the crystal to detector distance based on the diffraction limit of the crystal, the probable number of images to be collected based on the symmetry of the space group and the angle at which the crystal is rotated which depends on the size of the unit cell (Dauter, 1997). Hence, theoretically in order to collect most reflections for P222 space group, a 90° rotation would suffice. However, more images should be collected to achieve good redundancy in order to reduce errors. The higher the symmetry, the fewer degrees of rotation and therefore less number of images need to be collected in order to obtain a complete data set.

1.5.3 Data processing and scaling

After data collection, the raw intensities must be processed to improve their consistency and to maximize the number of measurements that are sufficiently accurate to be used. The procedure of processing the data involves the indexing of all of the reflections from planes in the chosen unit cell. This is followed by scaling of the data so that the backgrounds from different diffraction images match (Leslie, 2001). There are several software packages that automate this process and the most commonly used packages are the HKL package (Otwinowski and Minor, 1997)and MOSFLM (Leslie, 2001). The HKL package consists of: (i) DENZO which carries out autoindexing, refinement and integration (ii) XdisplayF, which functions to visualize each image/diffraction pattern, (iii) SCALEPACK, which is used for data scaling. Autoindexing using DENZO involves the determination of the standard lattice. A complete search of all possible indices of all reflections is performed. When the program finds values (integers) of one index (for example, h) for all reflections, this is equivalent to having found one real-space direction of the crystal axis (for example, a) and this is carried out by Fast Fourier Transform (FFT). After the search for real space vectors is completed, the program finds the three linearly independent vectors with minimal determinant (unit cell volume) that would index

all of the observed peaks and reduces the cell by conversion into a standard cell according to the Tables for X-ray Crystallography in the International Union of Crystallography, Vol I (Henry and Lonsdale, 1952), which contains an index for standard space group and symmetry classification. The space group of choice would usually be the one that has highest symmetry with least percentage of distortion and processing of the remainder of the data proceeds using the initial estimates as a reference. The next step to data processing involves averaging and merging the processed data by taking into account the errors which are caused by crystal decay and crystal absorption. The program SCALEPACK from the HKL package suite functions to serve this purpose. The output of this scaling process is a list of reflections with systematic absences characteristic of the space group that had been chosen during indexing (.sca file). The quality of the processed and scaled data is assessed by three values and can be viewed in the scalepack log file:

(i) Percentage completeness: Higher percentage completeness is favoured as this indicates that there is more information held in the processed data.

(ii) R_{merge}: The mean error of an intensity measurement, compared to the mean intensity, averaged over the chosen group of reflections (Blow, 2007).

$$R_{merge}\ (I) = \frac{\Sigma_h \Sigma_i |(I_h - I_{hi})|}{\Sigma_h \Sigma_i I_{hi}} \times 100$$

Where I_h is the weighted mean measured intensity of the observations I_{hi} in which the intensities of the symmetry related reflections are compared, thus R_{merge} is the measure of the disagreement $(I_h - I_{hi})$.

(iii) *I/σI:* σI is the uncertainty in the measurement of the intensity of a reflection. The acceptable resolution limit for scaling data is usually where the value of *I/σI* is more than 2.0.

MOSFLM (Leslie, 2001) consists of a graphical user interface through which the process of data processing is carried out and MOSFLM is relatively more straightforward than DENZO as it does not use key word input. The end result of data processing with MOSFLM is a raw MTZ file of reflection indices with their intensities and standard deviations. The MTZ file is then put through the program SCALA of CCP4 program package to be scaled and merged. The HKL suite and MOSFLM use SCALEPACK and SCALA respectively to perform statistical assessment of scaled data.

1.6 Structure factor and the phase problem

1.6.1 Structure factor

The vector (amplitude and phase) which represents the overall scattering from a particular set of Bragg planes is termed the structure factor (**F**). The structure factors for the various points on the crystal lattice correspond to the fourier transform of the electron density within the unit cell and vice versa(Blundell and Johnson, 1976).

$$F(h,k,l) = V\int_{x=0}\int_{y=0}\int_{z=0} \rho(x,y,z).\exp[2\pi i(hx + ky + lz)].dxdydz \qquad (1)$$

A reflection Electron density

V represents the volume of the unit cell. It arises because ρ is a density expressed as electron per unit volume, while the structure factor F is normally expressed as electrons per unit cell. Three integrations are taken over 0 to 1 in each of x, y, z, in order to extend it over the whole unit cell. The phase factor is $2\pi i(hx + ky + lz)$

The inverse fourier transform is,

All reflections phase

$$\rho(x,y,z) = 1/V\Sigma_h\Sigma_k\Sigma_l \mid F(h,k,l)\mid \exp-[2\pi i(hx + ky + lz)- i\alpha(h,k,l) \quad (2)$$

Electron density Amplitude

at a point

The goal of a structural analysis is to obtain the distribution of atomic electron density in the unit cell (in practice the atomic positions) starting from the diffraction data. This is not a straightforward task because the experimental data only brings forward the amplitude, but not the phases of the structure factors. F_{hkl} is a vector quantity that comprises of the amplitude, $|F_{hkl}|$ and its phases, $\alpha(h,k,l)$ as shown in equation **(1)**. From this equation, it is known that the frequency of the wave and the amplitude of the wave, which is the squared root of the measured intensity, I_{hkl} of the reflection with indices hkl. The information which cannot be accessed from the experimental data, is the phase or the angle at which the X-ray is diffracted. This is known as the 'phase problem'(Blundell and Johnson, 1976). This problem can be solved using several different methods: (i) Molecular replacement (MR). by determining the correct orientation and position of a molecule in the unit cell using a previously solved structure as a 'search model'(Rossman, 1972) (ii) Isomorphous replacement (IR) provides indirect estimates of the protein phase angles by observing the interference effects of the intensities on scattered beams by a heavy atom marker (Blundell and Johnson, 1976). (iii) Anomalous scattering in which the scattering information of an atom whose absorption frequency is close to the wavelength of the source beam produces phase information Multi-wavelength anomalous dispersion, requires intensity measurements at several wavelengths, while resolved anomalous scattering requires intensity measurements at one wavelength (Hendrickson, 1991). Both methods (ii) and (iii) can be used to solve novel structures for which there are no homologous structures to be used for molecular replacement.

1.6.2 Molecular replacement

A commonly used approach to solve the phase problem is Molecular Replacement (MR). Successful MR depends on the availability of a sufficiently homologous model structure. That is, if the structure of a protein (search molecule) is known that is homologous to the crystallized protein (target molecule) with unknown structure, the former can be used as a model to calculate a starting set of phases that can be interactively refined (Giacovazzo, 2002). For that, the search molecule must be oriented and positioned in the unit cell of the target molecule in a way that maximum overlap of the models calculated diffraction pattern with the observed diffraction pattern is achieved. Thus, the related structure (phasing model) is used to obtain phase information after orienting the model in the unit cell of the new protein. There are two major steps involved in MR, which are rotation and translation. In the rotation step, the spatial orientation of the known and unknown molecule with respect to each other is determined, while in the next step the translation needed to superimpose the now correctly oriented molecule onto the other is calculated (Figure 1.7). Several factors can affect the success of this procedure. Firstly, the sequence identity between the search model and the actual molecule should be high. Secondly the quality and completeness of the data should be as high as possible. Other factors include the relative size of the search model compared to the unit cell and how the agreement between the model and experimental data are assessed.

Several programs for MR are available and they include, AMoRe (Automated Molecular Replacement) (Navaza, 1994), MOLREP (Vagin and Teplyakov, 1997) and Phaser (Read, 2001; Storoni *et al.*, 2004) which are part of the CCP4 program suite(CCP4, 1994). Nevertheless finding the solutions to rotation and translation functions is not always a straightforward matter. In some cases it is necessary to modify the model, for instance, by ignoring the side chains and deletions/insertions in the model and by systematically varying the resolution range of the X-ray data used in the search. With the increasing number of protein structure determinations, MR has become an increasingly popular technique for protein phase angle determination.

If more than one molecule is present in the asymmetric unit, molecular replacement would have to be repeated for each of the molecules. The knowledge of the number of molecules in the asymmetric unit can be determined from the Matthews coefficient (Matthews, 1968) which from the size of the protein and the unit cell estimates the solvent content of the crystal for different numbers of molecules in the unit cell. These solvent contents are then compared to the average solvent content of protein crystals to determine how many molecules are present in the asymmetric unit.

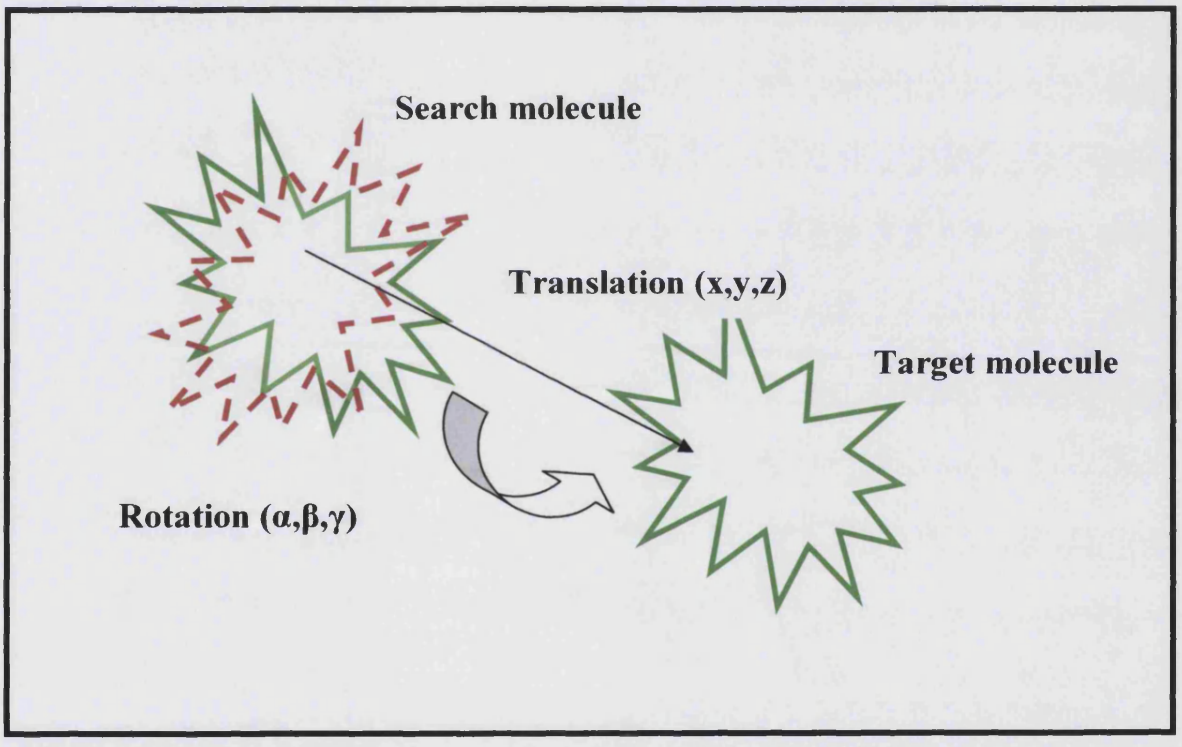


Figure 1.7: A schematic illustration of the molecular replacement method. The search molecule is represented by the figure in red. The target molecule is represented in green

1.6.3 Isomorphous replacement and anomalous scattering (Perutz, 1956).

The *Isomorphous replacement* method utilizes phase information from the comparison of structure factors from native and derivative crystals soaked in heavy metals. Because atoms contribute most strongly to structure factors from planes which intersect their positions, the uniform addition of heavy atoms, which have larger scattering factors, at certain points in the unit cell gives rise to an increase in the amplitude of those structure factors. If the native and the derivative crystals are isomorphous the resulting intensities can be compared to the native intensities to retrieve phase information for these atoms. This phase information can then be placed in the unit cell so that a rough framework for the model to be built into can be made. To determine the phases as accurately as possible, several derivatives are often necessary, requiring the production of many isomorphous derivative crystals. *Anomalous scattering* utilizes the absorption of X-rays by heavy atoms to derive information about phases (Hendrickson, 1991). All diffraction patterns are centrosymmetric as diffraction from each plane gives reflection, *hkl* and *–h-k-l* which are equal in intensity. This is known as Friedel's law. All atoms absorb radiation at specific wavelengths which correspond to the required energy for electron excitation. For certain heavy atoms, these wavelengths correspond to X-ray wavelengths suitable for crystallography. If the X-rays used are tuned to this absorption edge, the absorption of the radiation results in an observable breakdown in Friedel's law known as anomalous scattering, which can be used to retrieve phase information for these atoms. This technique has an advantage over isomorphous replacement methods as non-isomorphism of heavy atom soaks can be avoided by recording both native (at low absorbing wavelength) and anomalous datasets from the same crystal. Another advantage is with the option of using selenium that has an absorption edge of 0.98 Å, which can be incorporated during expression of the protein. Bacterial cultures that are grown in seleno-methionine media can be used to produce recombinant protein with methionine uniformly replaced with seleno-methionine, hence preventing the need for crystal soaking.

The use of a combination of both these methods to estimate phases is also becoming increasingly common. Phase information from the increased scattering factor and the anomalous scattering of heavy atoms can be combined in a technique referred to as *Single Isomorphous Replacement with Anomalous Scattering* (SIRAS).

1.7 Model building, crystallographic refinement, structure validation, structure analysis and deposition

1.7.1 Model building

When a rotation and translation operation for MR, or the heavy atom positions have been found, the model phases can now be incorporated with the observed amplitudes from the diffraction pattern and the structure factors can then be refined against the model structure. If the model phases are correct, the R-factor which is the ratio of the difference between computed amplitudes for the model and the observed amplitudes should decrease. These structure factors can then be used to calculate an initial map which if the phase information is good enough should be interpretable. Crystallographic refinement is carried out in a cyclic process of gradual improvement of the model. During the refinement process the model is changed during model building at the graphics display in such a way that the structure factors calculated from the built model get closer to the observed amplitudes. Iterative cycles of refinement optimize the atomic model, allow for improved phases and thus help compute electron density maps (refer to equation (2), page 21) that show finer details of the molecular structure. The cyclic process of model building and refinement is usually repeated until ultimately a model which represents the observed data as closely as possible is generated.

Model building requires a certain degree of map interpretation skills. The electron density equation generates values for electron density for each point, $\rho(x,y,z)$, in the unit cell, from the structure factors. These figures are then displayed as three dimensional contour maps on a graphic computer interface using programs such as O(Jones *et al.*, 1991) or COOT(Emsley, 2004). In order not to bias the map

to the model structure-factor amplitudes, these are usually subtracted from a multiple of the observed amplitudes for example; a $2F_o$-F_c map is the sum of twice the observed structure factor amplitudes, minus calculated amplitudes, multiplied by the phase term. This map is used to represent the positive continuous density of the model while F_o-F_c maps are generated in order to provide positive peaks which, indicates that there is something in the data which is lacking in the model and it usually helps in identifying solvent molecule as well as ligands and ions that are either an integral part of the macromolecule or have been incorporated into the crystal lattice from the crystallisation buffer. Negative peaks in FoFc maps which indicate there is something in the model which is not present in the data. The respective maps are usually contoured to a set sigma level for viewing in the graphics although local adjustment of these levels is often necessary to assist in map interpretation. The quality of the map depends on the accuracy or inaccuracy of the phases and is also dependent on the resolution.

1.7.2 Refinement

As mentioned before, between rounds of model building, refinement is also carried out in order to improve the model. In general, refinement can be described as the optimization of a function, to find the global minimum, by changing the parameters of the model. The model parameters include the position, occupancy and thermal factors (B-factors) of the atoms. The algorithm used for refinement varies between refinement programs. There are several programs that are available for molecular refinement purposes. SHELX (Sheldrick *et al.*, 1993) utilizes the least squares (LS) function which minimise the sum of the weighted (Murshudov *et al.*, 1997)difference between two sets of observations (the observed and the calculated amplitudes). An alternative to the least square algorithm is the Maximum likelihood (ML) function which is incorporated in REFMAC (Murshudov *et al.*, 1997) and CNS (Brunger *et al.*, 1998). The ML function finds the set of parameters which are the most probable to produce the experimental results seen. Another method which

is also incorporated in CNS is energy minimization for which the target is the model with the lowest calculated energy.

In order to avoid model bias and entrapment in local minima, it is often advisable to initially start refinement at a lower resolution and then subsequently increase the resolution in one or more steps. This ensures that the gross features of the model are corrected first and the wrongly assigned details do not bias the model. Using restraints and constraints in refinement is also important especially at lower resolution. Adding restraints increases the observations to parameter ratio hence, a good technique is to restrain the geometry of the protein tightly so as the phases could become more accurate as distortions in local geometry cannot be assigned without good phases. Restraining of non-crystallographic symmetry partners together is also of particular use as it enables the averaging out of noise in the phases. Initial refinement is usually rigid body refinement in which the protein is refined as a whole followed by optimization of atom positions. If the data are of reasonably high resolution (better than 3 Å), individual water molecule positions and protein atoms B factors can also be refined. B-factors are atomic displacement factors and are very closely associated with atomic coordinates of the model. B-factors represent the distribution of positions occupied by an atom over a period of time (dynamic disorder) as well as variations in the position of an atom between different unit cells (static disorder). Large B-factor values are usually indicative of errors in the model coordinates.

1.7.3 Structure validation

The measure of the success of a refinement process can be assessed by several means:

(i) R-factor: The primary measure of a refining structure is given by the residual index or the R-factor. As the model improves, the measured Structure Factors (F_o) and the calculated (F_c) begin to converge and the value of the R-factor drops:

$$R_{cryst} = \frac{\Sigma \, ||F_o| - |F_c||}{\Sigma |F_o|}$$

F_0 is derived from the measured reflection intensity and F_c is the corresponding structure factor obtained from the current model. This conventional parameter is known as the crystallographic reliability index (R_{cryst}).

(ii) Rfree: R_{free} was first developed by Brunger and co-workers in 1997. R_{free}(Brunger *et al.*, 1998) is a more objective indicator of of model quality than R_{cryst}. Rfree is calculated as the Rcryst but for a randomly selected percentage of reflections known as the test set which are not used in the refinement. This acts as a cross validation tool to check that the refinement is not biasing the model away from the data

(iii) Ramachandran plot (Ramachandran *et al.*, 1963): The plot highlights the stereochemistry, conformation (planarity of the peptide bonds, bond lengths, bond angles etc.) and the backbone conformation angles (Φ,Ψ) of the current model based on the expected values for simple organic molecules. The program PROCHECK (CCP4, 1994) is a useful program which can assist in the assessment of the quality of the structure at various stages of refinement.

(iv) Root mean square deviation (r.m.s.d): r.m.s.d value is another statistical method that helps assess the quality of a structure. The value measures the deviation from ideal geometry and a low value reflects on the success of refinement. Well refined structures would have a value between 0.005-0.010.

1.7.4 Structure analysis and PDB deposition

Once the structure is complete, the next step is to deposit the coordinates with the Protein Data Bank (PDB) at either European Bioinformatics Institute (EBI; http://www.ebi.ac.uk/) or with the Research Collaboratory for Structural Bioinformatics (RCSB; http://www.rcsb.org/). The PDB helps in providing access to structural information (for proteins, nucleic acids, ligands and other biological macromolecules) to scientists all over the world. Studying the three-dimensional structure of a protein can bring insight into the functional aspects of the protein. Therefore, analytical observations should be carried out on the structural data obtained. The structure of a refined model does not only reveal the detailed architecture of the polypeptide chain folds but also gives an insight into how proteins evolved, the basis on which they interact with each other and how they bind with other molecules. Several programs are available to analyse structures and some of these are DSSP (Kabsch and Sander, 1983) which functions to assign secondary structure elements in the structure, CONTACT (CCP4, 1994) which is useful for analyzing hydrogen-bonding interactions, SC (CCP4, 1994) for analysis of the shape complimentarity between accessible area in a protein structure, CAVENV (CCP4, 1994) to visualize cavities in the protein structure and PROSA (Sippl, 1993) which functions to determine the native structural fold of the protein.

CHAPTER 2

Glycosyltransferases

2.1 General overview

The study of the biosynthesis of complex carbohydrates and polysaccharides is becoming increasingly popular in the field of glycobiology. This is largely due to the important roles that complex carbohydrates play in cellular systems. These molecules govern a diverse range of cellular functions, including energy storage, cell wall structure, cell-cell interactions and signalling, host-pathogen interactions, and protein glycosylation (Bertozzi and Kiessling, 2001; Cautinho *et al.*, 2003; Rudd, 2001; Wells, 2001). In these functions especially where these carbohydrates act as a cellular language, they rely on the precise carbohydrate structures which display large chemical diversity. The biosynthesis of these carbohydrates may involve the action of a wide range of different and selective glycosyltransferases, enzymes which function to transfer sugar moieties from activated donor substrates to specific acceptor substrates. Even though many glycosyltransferases (GTs) catalyse chemically similar reactions, they display remarkable diversity in their donor, acceptor and product specificity hence, are able to generate a potentially infinite number of glycoconjugates, oligo- and polysaccharides.

Transfer of the sugar residue occurs with either the retention or the inversion of the configuration of the anomeric carbon (Figure 2.0) (Breton *et al.*, 2006). These enzymes can be found in both eukaryotes and prokaryotes and in eukaryotes, most of the glycosylation reactions occur in the Golgi apparatus. These eukaryotic GTs (Gibson *et al.*, 2002) belong to the type-II transmembrane proteins which have a large C-terminal globular catalytic domain facing the luminal side. The recent elucidation of several crystal structures of GTs, have brought about greater understanding of the molecular basis of mechanism and specificity for these enzymes creating tremendous possibility for inhibitor studies and facilitating the engineering of new catalysts for the enzymatic synthesis of natural and novel glycoconjugates for therapeutic purpose. Glycosyltransferases using nucleotide diphospho-sugars, nucleotide monophosphate sugars, sugar phosphates and related proteins are classified into families based on the amino-acid sequence similarities developed by Coutinho (Coutinho *et al.*, 2003a). At the time of writing (May 2007),

the Carbohydrate Active Enzymes (CAZy) database (http://afmb.cnrs-mrs.fr/CAZY) comprises more than 89 families of GTs. Large differences are observed in the number and function of families of GTs. Several families comprise of a huge number of sequences from various sources with diverse functions. This is best illustrated by the family 2 GT which contains more than 3000 sequences originating from various species of animal, plant, yeast and bacteria and for which 12 distinct functions have already been identified. Some of the members of this family include cellulose synthase, chitin synthase, mannosyltransferase, glucosyltransferase and galactosyltransferase. In the case of the polyspecific GT such as those from family GT2 and GT4, sequence similarities are restricted to only a portion of the catalytic domain while the monospecific family GT will generally display sequence similarity for the whole catalytic domain.

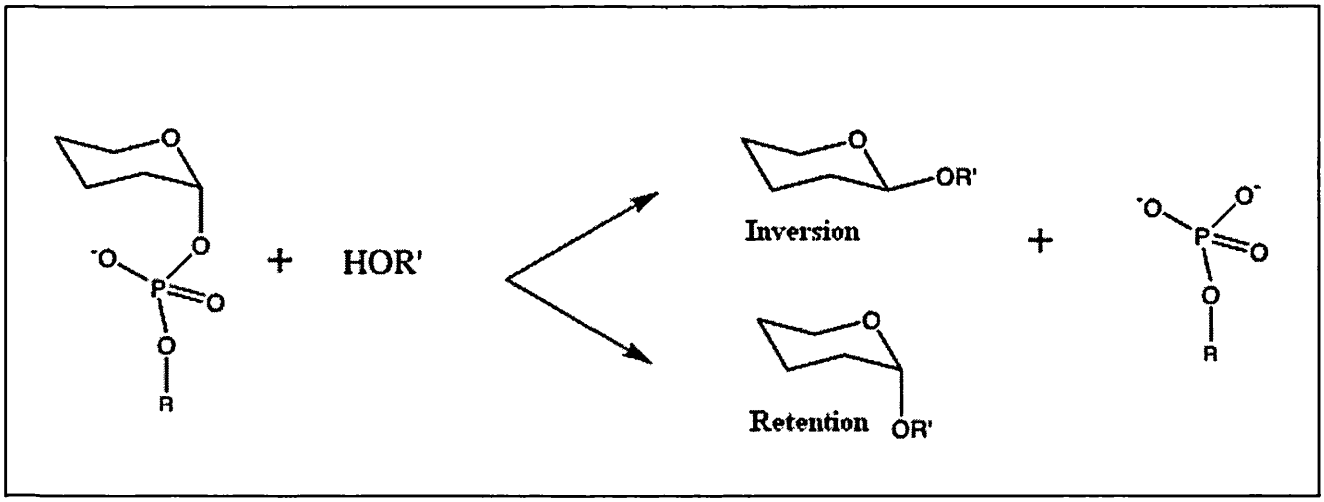


Figure 2.0: Glycosyltransferases use sugar donors which typically have an activating group (substituted) phosphate such as a nucleotide (R=nucleoside monophosphate) or lipid phosphate (R=lipid). The acceptor is shown as R'OH where R can be varied, i.e, a sugar, lipid, protein, antibiotic or nucleic acid. Catalysis may occur with two possible outcomes; inversion or retention of the anomeric configuration of the donor (Coutinho *et al.*, 2003b).

These enzymes also display a range of diversity in their catalytic activity despite having remarkably high sequence identity or similarity making it difficult to predict the function of a putative GT of sequence similarity. This is reflected in the study of the A and B blood group enzymes, which showed different donor substrate specificity [A transferase, UDP-N-acetylgalactosamine (UDP-GalNAc) and B transferase, UDP-galactose (UDP-gal)] despite differing by only four amino acids (Yamamoto *et al.*, 1990). The classification system which is currently being used for GT is based on sequence similarities, however this mode of classification may not apply to all GT families as distant similarities between inverting and retaining families have been identified which suggests that these families may share a common ancestor (Franco and Rigden, 2003; Liu and Mushegian, 2003). The recent elucidation of the crystal structure of mannosylglycerate synthase (MGS) from *Rhodobacter marinus* showed that although at the sequence level the protein showed similarity to the GT2 family, it is a retaining enzyme that has been classified into a new family (GT78) (Flint *et al.*, 2005). This shows that the mechanism of GT cannot be predicted with reliability based on sequence similarities alone; hence the classification system should also integrate both structural and mechanistic features within each family.

The elucidation of the crystal structures of GT enzymes is slowed down by the difficulties in high-level expression, purification and crystallisation, however, even with this major hampering factor, more than 100 crystal structures representing 39 different GT enzymes from both prokaryotes and eukaryotes have been elucidated thus far (Table 2.0). GT structures have been generally observed to adopt a α/β/α sandwich (Figure 2.1) which is very closely similar to the Rossman-type fold found in many nucleotide binding proteins (Lesk, 1995). Originally GTs could be structurally divided into 2 superfamilies, named GT-A and GT-B which were first observed in the SpsA and β-glucosyltransferase (BGT) structures (Charnock and Davies, 1999b; Vrielink *et al.*, 1994). The general characteristics of a GT-A structure comprises of α/β proteins with a single Rossman domain, conserved 'DXD' motif and a conical active site cleft formed by two associated domains (Charnock and Davies, 1999a; Persson *et al.*, 2001; Unligil *et al.*, 2000) while the

GT-B group consists of two Rossmann domains separated by a deep substrate-binding cleft (Gibson *et al.*, 2002; Ha *et al.*, 2000; Vrielink *et al.*, 1994). Recently, a novel fold which is closely similar to the GT-A fold but with some differences has been reported for the bacterial sialyltransferase (CstII) which belongs to family GT42 (Chiu *et al.*, 2004) (Figure 2.1). This particular structure exhibits a single Rossman domain hence falls under the GT-A superfamily, however, marked differences from other member of the GT-A family are also noted in terms of the connectivity of the secondary structural elements and the absence of the conserved DXD motif. The GT-A and GT-B fold are not exclusively found only in GTs but they can also be found in non-GT enzymes such as nucleotidyltransferases and sugar epimerases. This chapter will discuss in further detail the structural classification of GTs, the molecular dynamics that these enzyme undergo during catalysis and their catalytic mechanism.

Organism	Glycosyltransferase	Name	GT Family	Structural superfamily	Reference
Virus					
Phage T4	β-Glucosyltransferase	BGT	GT63	GT-B	(Vrielink *et al.*, 1994)
Phage T4	α-Glucosyltransferase	AGT	GT72	GT-B	(Lariviere *et al.*, 2005)
Prokaryotes					
Agrobacterium tumefaciens	Glycogen synthase I	AtGS	GT5	GT-B	(Buschiazzo *et al.*, 2004)
Amycolatopsis orientalis	β-Epi-vancosaminyltransferase	GtfA	GT1	GT-B	(Mulichak *et al.*, 2003)
	β-Glucosyltransferase	GtfB	GT1	GT-B	(Mulichak *et al.*, 2001)
	β-Vancosaminyltransferase	GtfD	GT1	GT-B	(Mulichak *et al.*, 2004)
Bacillus subtilis	Putative glycosyltransferase	SpsA	GT2	GT-A	(Charnock and Davies, 1999b)
Campylobacter jejuni	α-2,3/2,8-Sialyltransferase	CstII	GT42	Novel, similar to GT-A	(Chiu *et al.*, 2004)
Escherichia coli	β-1,4-GlcNAc transferase	MurG	GT28	GT-B	(Ha *et al.*, 2000)
	Trehalose-6-phosphate synthase	OtsA	GT20	GT-B	(Gibson *et al.*, 2002)
	Heptosyltransferase II	RfaF	GT9	GT-B	PDB code: 1PSW
Neissseria meningitidis	α-1,4-Galactosyltransferase	LgtC	GT8	GT-A	(Persson *et al.*, 2001)
Rhodothermus marinus	Mannosylglycerate synthase	MGS	GT78	GT-A	(Flint *et al.*, 2005)
Streptomyces antibioticus	UDP-Glucosyltransferase	OleI/OleD	GT1	GT-B	(Bolam *et al.*, 2007)
Pyrococcus abyssi	glycogen synthase	PaGS	GT5	GT-B	(Horcajada *et al.*, 2006)
Clostridium difficile	α-Glucosyltransferase	Toxin B	GT44	GT-A	(Reinert *et al.*, 2005)
Mycobacterium smegmatis	Phosphatidylinositol Mannosyltransferase	PimA	GT4	GT-B	(Guerin, 2007)
Streptococcus pneumoniae	penicillin-binding protein 1B	Pbp1b	GT51	GT-A	(Lovering *et al.*, 2006)
Aquifex Aeolicus	Peptidoglycan glycosyltransferase	MrcA	GT51	GT-A	(Yuan *et al.*, 2007)
Pasteurella multocida	α-2,3-sialyltransferase	PmST1	GT80	GT-B	(Ni *et al.*, 2006)
Streptomyces viridochromogenes	Eurekanate-attachment enzyme	AviGT4	GT4	GT-B	(Martinez-Fleites *et al.*, 2006)
Eukaryotes					
Yeast	α-1,2-Mannosyltransferase	Kre2p/Mnt1P	GT15	GT-A	(Lobsanov *et al.*, 2004)
Mouse	α-1,4-N-Acetylhexosaminyltransferase	Ext12	GT64	GT-A	(Pedersen *et al.*, 2003)
	Polypeptide-α-GalNAc transferase	ppGalNAc-T1	GT27	GT-A	(Fritz *et al.*, 2004)
	β-1,3-N-Acetylglucosaminyltransferase	Mfng	GT31	GT-A	(Jinek *et al.*, 2006)
	β-1,6-N-acetylglucosaminyltransferase	C2GNT	GT14	GT-A	(Pak *et al.*, 2006)
Rabbit	α-Glucosyltransferase	Glycogenin	GT8	GT-A	(Gibbons *et al.*, 2002)
	β-1,2-GlcNAc transferase I	GnTI	GT13	GT-A	(Unligil *et al.*, 2000)

Organism	Glycosyltransferase	Name	GT Family	Structural superfamily	Reference
	α-1,3-mannosyl-glycoproteinβ-1,2-N-acetylglucosaminyltransferase I	Mgat1	GT13	GT-A	(Gordon *et al.*, 2006)
Bovine	α-1,3-Galactosyltransferase	α3GT	GT6	GT-A	(Gastinel *et al.*, 2001)
	β-1,4-Galactosyltransferase I	β4Gal TI	GT7	GT-A	(Gastinel *et al.*, 1999)
Human	β-1,3-Glucuronyltransferase	GlcAT-I	GT43	GT-A	(Pedersen *et al.*, 2000)
	β-1,3-Glucuronyltransferase	GlcAT-P	GT43	GT-A	(Kakuda *et al.*, 2004)
	α-1,3-GalNAc transferase A	GTA	GT6	GT-A	(Patenaude *et al.*, 2002)
	α-1,3-Galactosyltransferase B	GTB	GT6	GT-A	(Patenaude *et al.*, 2002)
	β-1,4-Galactosyltransferase	B4Gal-T1	GT7	GT-A	(Ramakrishnan *et al.*, 2006)
Vitis vinifera	Anthocyanidin-3-O-Glc transferase	VvGT1	GT1	GT-B	(Offen *et al.*, 2006)
Medicago truncatula	β-Glucosyltransferase	UGT71G1	GT1	GT-B	(Shao *et al.*, 2005)

Table 1: A list of crystal structures of different types of glycosyltransferases. All data are available from the 3D-Glycosyltransferase database (http://www.cermav.cnrs.fr/glyco3d/). Table adapted from (Breton *et al.*, 2006)

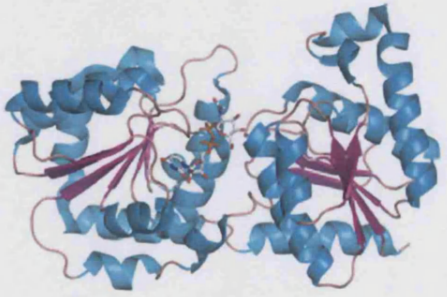
2.1.1 Structural classification of Glycosyltransferase

As previously mentioned, the GT-A fold resembles a Rossmann fold consisting of an α/β/α sandwich where the β-sheet is a seven stranded β-strand with strand 6 running antiparallel to the rest. This central β-sheet is flanked by a smaller β-sheet which forms the active site of the enzyme. Another general feature of the GT-A enzymes is that they possess a common motif, the DXD motif which is associated with their dependence on a divalent cation for catalytic activity (Breton *et al.*, 1998; Breton and Imberty, 1999; Breton *et al.*, 2006). In the structures of GTs belonging to the GT-A family, this DXD motif has been shown to interact with the diphosphate region of the nucleotide donor through the coordination of a divalent metal cation, which in most cases is Manganese. The two aspartate amino acids in the DXD motif are not always conserved across the GT-A family but this particular motif or its variants can always be identified at the same location, in a short loop connecting the central β-sheet of the α/β/α sandwich to a smaller β-sheet (Figure 2.1 A). There are two regions which are structurally well conserved in all members of the GT-A family, including both inverting and retaining enzymes, which suggests that common structural elements are necessary for the glycosyl transfer reaction, irrespective of the stereochemistry of the reaction. The first region corresponds to the Rossmann-type nucleotide binding domain encompassing the first 100-120 residues and terminating at the DXD motif. The key amino acids which interact with UDP are mainly located between β1 to β5 strands. In some crystal structures, residues in the C-terminal region were also shown to make interactions with UDP (Boix *et al.*, 2001) (Figure 2.3). Differences are also observed in the function of the DXD motif residues in retaining and inverting enzymes. In the retaining enzymes, the two aspartate residues can interact with the manganese ion, whereas only the last aspartate interacts with the metal cation in inverting enzymes (Breton *et al.*, 2006; Persson *et al.*, 2001; Tarbouriech *et al.*, 2001). In both cases the DXD motif plays a crucial part in donor nucleotide binding by mediating the correct coordinating interactions for optimal binding of the donor molecule. The second structurally conserved region is the β-α-α region which has been shown in many crystal

structures of GTs to interact with both the donor sugar and the sugar acceptor (Boix *et al.*, 2002; Breton *et al.*, 2006; Persson *et al.*, 2001). Despite the similarity of their spatial folds, GT-A and GT-B enzymes appear to be unrelated. Both families have members in the three domains of life and in addition, some bifunctional enzymes are expected to contain both GT-A and GT-B folds.

Figure 2.1: Structural representatives of the 3 different folds displayed by GTs thus far. Bound nucleotide is shown as sticks while manganese ion when present is shown as a sphere. (A) GT-A fold, mouse α-1,4-*N*-acetylhexosaminyltransferase (EXTL2) complexed with UDP-GalNAc. The conserved DXD motif is highlighted in green (PDB code: 1OMZ) (Pedersen *et al.*, 2003). (B) GT-B fold, *Escherichia coli* MurG complexed with UDP-GlcNAc (PDB code: 1NLM) (Hu *et al.*, 2003). (C) *Campylobacter jejuni* sialytransferase CstII complexed with cytidine monophospho 3-fluoro N-acetyl neuraminic acid (CMP-3FNeuAc) (PDB code: 1RO7) (Chiu *et al.*, 2004) (Adapted from (Breton *et al.*, 2006).

GT-A



GT-B

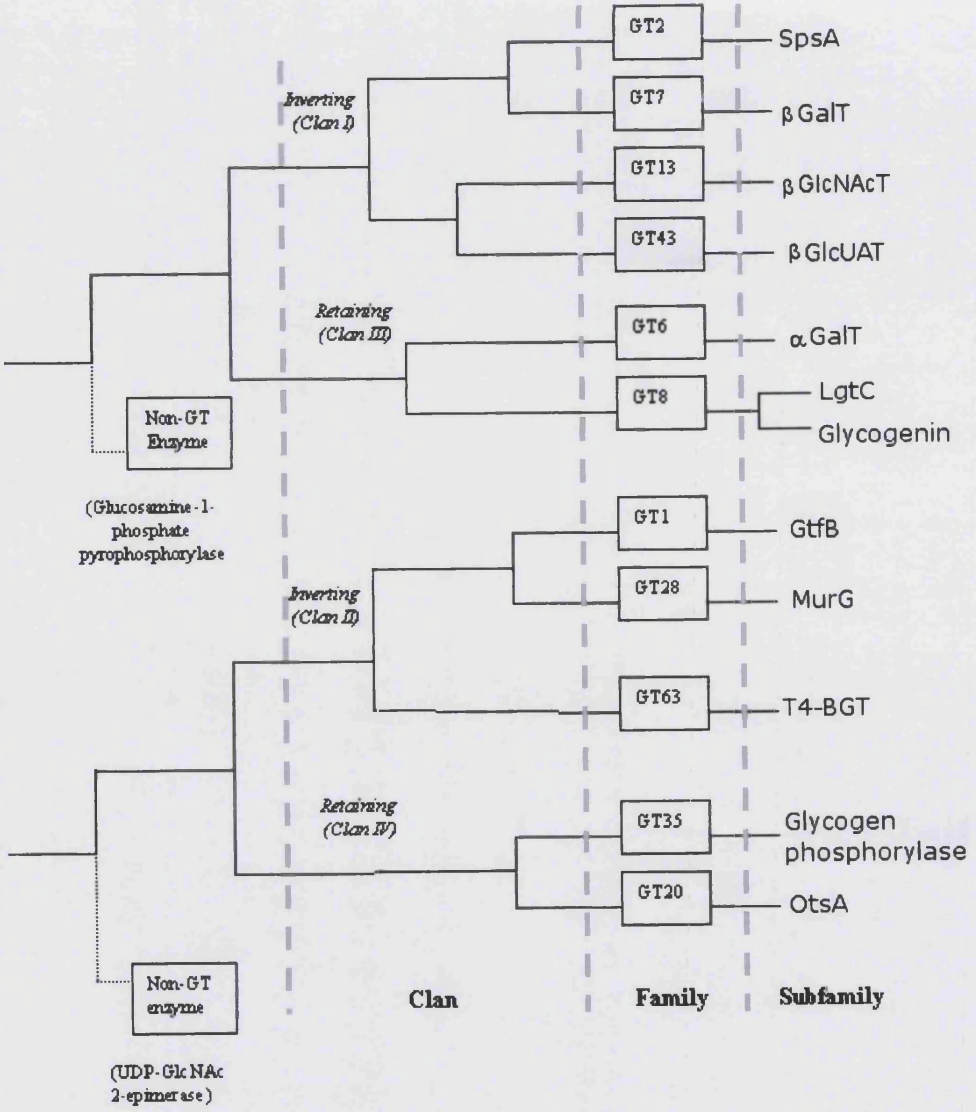

GT2
SpsA
GT7
β GalT
GT13
β GlcNAcT
GT43
β GlcUAT
Inverting (Clan I)
Retaining (Clan III)
GT6
α GalT
GT8
LgtC
Glycogenin
Non-GT Enzyme
(Glucosamine-1-phosphate pyrophosphorylase
GT1
GtfB
GT28
MurG
Inverting (Clan II)
GT63
T4-BGT
Retaining (Clan IV)
GT35
Glycogen phosphorylase
GT20
OtsA
Non-GT enzyme
(UDP-GlcNAc 2-epimerase)
Clan
Family
Subfamily

Figure 2.2: The hierarchical classification of glycosyltransferases from folds to clans, families and subfamilies, illustrated for some of the glycosyltransferases with a reported 3-D structure. SpsA, Bacillus subtilis glycosyltransferase (Charnock and Davies, 1999b) βGalT, bovine β-1,4 galactosyltransferase (Gastinel *et al.*, 1999) βGlcNAcT, rabbit β-1,2-N-acetylglucosaminyltransferase I (Unligil *et al.*, 2000)βGlcUAT, human β-1,3 glucuronyltransferase (Pedersen *et al.*, 2000) αGalT, bovine α-1,3-galactosyltransferase (Gastinel *et al.*, 2001) LgtC, Neisseria meningitides α-galactosyltransferase (Persson *et al.*, 2001)glycogenin, rabbit glycogenin (Gibbons *et al.*, 2002)GtfB, Amycolatopsis orientalis β-glucosyltransferase (Mulichak *et al.*, 2001) MurG, *Escherichia coli* β-N acetylglucosaminyltransferase (Ha *et al.*, 2000) T4-BGT, bacteriophage T4 DNA β-glucosyltransferase (Vrielink *et al.*, 1994) glycogen phosphorylase, rabbit glycogen phosphorylase (Goldsmith *et al.*, 1989) OtsA, E. coli trehalose-6-phosphate synthase (Gibson *et al.*, 2002). Adapted from (Coutinho *et al.*, 2003a).

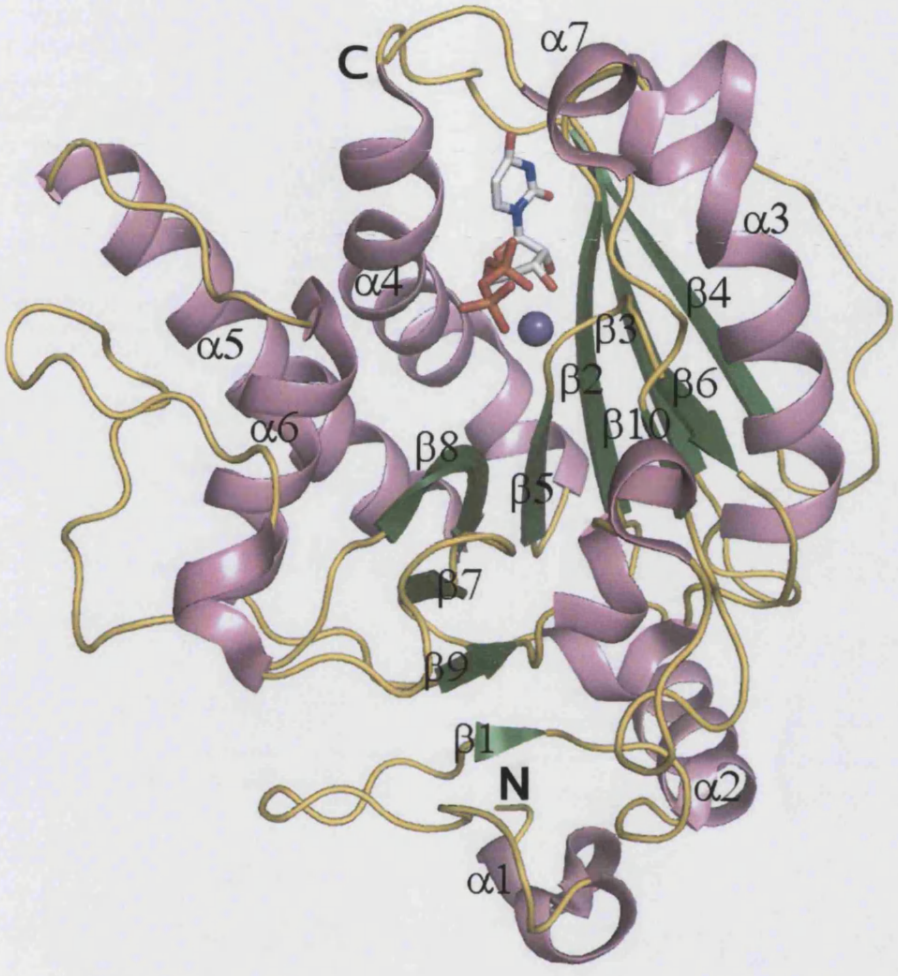


Figure 2.3: Typical GT-A fold shown here by the crystal structure of bovine α-1,3 galactosyltransferase (α3GT) in complex with manganese ion and UDP. The bound ligand and ion identify the location of the active site. The manganese ion is shown as a purple sphere, UDP is in white, and helices are pink, while the strands are green [Adapted from (Boix *et al.*, 2001)].

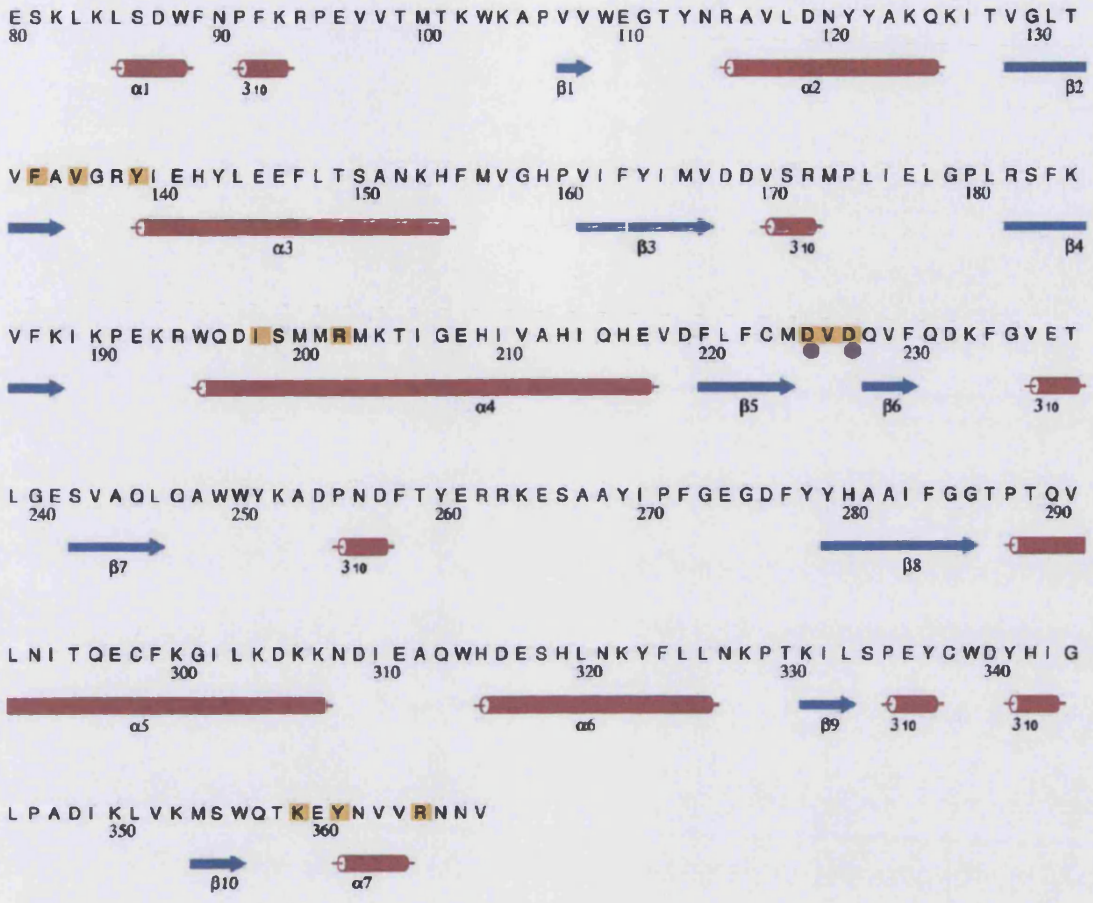


Figure 2.4: The amino acid sequence of the catalytic domain of α3GT with all secondary structure elements highlighted. UDP binding residues are marked in yellow, while the manganese binding residues are shown by closed magenta spheres. [Adapted from (Boix *et al.*, 2001)]

A novel fold family has recently been reported with the determination of the crystal structure of sialyltransferase (CstII) from *Campylobacter jejuni* (Chiu *et al.*, 2004). Although this enzyme showed structural similarities to the GT-A type fold it also displayed several differences as it has a different type of α/β/α sandwich (a seven stranded parallel β-sheet) and does not have a DXD motif.

2.1.2 Conformational dynamics of glycosyltransferases

The elucidation of several crystal structures of GTs in complex with substrates or analogues have shown that these enzymes undergo large conformational movements during the catalytic reaction of which a major part occurs due to the binding of the respective donor and acceptor sugars. The ratio of loops to secondary elements is high in GTs and many of the crystal structures which have been solved thus far do not describe the entire catalytic domain because some of the loops present in the enzyme are highly flexible and do not present distinct electron density. In GT, these flexible loops appear to play an important role in substrate binding as there have been demonstrations in some of these enzymes of an ordered binding of donor and acceptor substrates linked to a donor substrate conformational change (Boix *et al.*, 2002; Boix *et al.*, 2001; Ramakrishnan and Qasba, 2002). In several of these structures, a disordered loop or the C-terminal region becomes ordered on binding of the nucleotide donor to create a lid which is folded over the donor substrate making direct contacts with the diphosphate region of the donor nucleotide. This conformational change results in the formation of a closed active site pocket that will serve as a binding site for the acceptor molecule. The structure of rabbit β-1,2-*N*-acetylglucosaminyltransferase I (GnT I) contains a 13-residue internal stretch that undergoes loop ordering upon binding of the donor nucleotide, UDP-*N*-acetylglucosamine (Unligil *et al.*, 2000). This conformational change was postulated to result in the formation of one side of a pocket surrounding the active site, permitting catalysis to occur. The β-glucosyltransferase from the T4 phage is another enzyme which displays ordering of a disordered loop upon binding of a donor substrate. This enzyme in its resting state has two disordered loops in its N-terminal domain; however in the presence of UDP-glucose donor, both these loops become ordered (Vrielink *et al.*, 1994) and similarly, the structure of LgtC galactosyltransferase from *Neisseria meningitidis* also shows two loops that undergo stabilisation on binding of donor galactose. The stabilised loops were shown to form several intramolecular van der Waals contacts with the donor molecule (Persson *et al.*, 2001). A series of amino acids in β-1,4-galactosyltransferase (β4Gal-T1)

undergoes a conformational change upon binding of the donor substrate (Gastinel *et al.*, 1999; Ramakrishnan and Qasba, 2001). In the absence of α-Lactalbumin (LA), β4Gal-T transfers galactose to N-acetylglucosamine but in the presence of LA the acceptor specificity is changed to glucose. This combination of β4Gal-T and LA is called lactose synthase (LS). β4Gal-T on its own has a short loop of 8 amino acids but in presence of LA, a much larger loop of 21 amino acids that incorporates the smaller 8 amino acids loop is formed. The binding of the donor molecule to this LS complex results in the reorientation of the 21 amino acid loop such that 14 amino acids adopt a loop configuration whereas the remaining 7 residues form an α-helix (Yazer and Palcic, 2005). This results in the movement of His347 into position to coordinate with manganese and covers the bound substrate for catalysis to take place (Ramakrishnan and Qasba, 2001; Yazer and Palcic, 2005). This conformational change is thought to prevent water molecules from acting as an acceptor for sugar transfer hence, minimising hydrolysis of the energetically precious nucleotide sugar. The mouse β-1,3-N-acetylglucosaminyltransferase structure in complex with UDP showed disorder in two loop regions and based on mutational studies, one of the longer loop residues was predicted to be the putative site for acceptor molecule binding (Jinek *et al.*, 2006). However, the human glycosyltransferases, GTA and GTB do not display reordering of their 26 amino acids disordered loop upon binding of their respective donor molecules, suggesting that these residues might already be in a catalytically active position independent of the binding of the donor or else are not involved in determining donor or acceptor specificity (Patenaude *et al.*, 2002; Yazer and Palcic, 2005). Mutational studies of residues on the flexible loops of GTs serve to strengthen the hypothesis that these loops may play a part in catalysis or at the very least in the binding of both donor and acceptor molecules. A point mutation of Gly320Asp which is located on a disordered loop in the GnT 1 rendered the enzyme completely inactive (Unligil *et al.*, 2000) while mutation of Arg365 on the C-terminus of α3GT to Lysine significantly decreased the catalytic activity of the enzyme by possibly compromising the stability of the transition state in the reaction (Boix *et al.*, 2001). In addition, affinity studies using titration microcalorimetry on LgtC (GT8) demonstrated that the open state (free enzyme) has very little affinity

for the oligosaccharide acceptor. Since both inverting and retaining glycosyltransferases undergo conformational changes during catalysis (Boix *et al.*, 2002; Boix *et al.*, 2001; Ramakrishnan and Qasba, 2001), there is an increasing interest towards the study of conformational dynamics in these enzymes both by structural means and also by means of molecular dynamic simulations (Gunasekaran and Nussinov, 2004; Snajdrova *et al.*, 2004).

2.1.3 Mechanism of glycosyltransferase

A specific feature of GTs is that they use an activated donor that can either be a nucleoside diphosphosugar, nucleoside monophosphosugar or lipid phosphosugar. Mechanistically, GTs can be divided into two groups, the inverting GTs and the retaining GTs. During GT catalysis, two stereochemical outcomes are possible for the hydrolysis of a glycosidic bond: the anomeric configuration of the product can either be retained or inverted with respect to the starting material. The mechanism for inverting GTs is believed to be similar to that of inverting glycosylhydrolases with the requirement of one acidic amino acid that activates the acceptor hydroxyl by deprotonation (Figure 2.5a). All structural data so far lend support to the S_N2 mechanism first proposed by Murray *et al.* for α3-FucT (Murray *et al.*, 1996). Inverting GTs which have the GT-A fold use a divalent cation , generally manganese in catalysis where this ion plays a role of acid catalyst and also initiates a sequential ordered mechanism, in which nucleotide sugar binding is followed by restructuring of a loop region to form a closed active site hence, creating the binding site for the acceptor molecule (Boix *et al.*, 2002; Ramakrishnan *et al.*, 2004). GTs which possess the GT-B fold on the other hand do not need a divalent metal cation for their catalysis. T4 phage BGT which transfers glucose to modified DNA also uses an S_N2 mechanism with the participation of one acidic amino acid from the protein (Lariviere *et al.*, 2005).

Thus far, the mechanism of retaining GTs has yet to be established. Direct comparison with glycosylhydrolases would suggest a double displacement mechanism which involves the formation of a short-lived glycosyl-enzyme

intermediate (Figure 2.5b). If this mechanism is to take place, it demands the existence of an appropriately positioned nucleophile within the active site. As in the inverting enzyme mechanism, a divalent metal cation will play the role of a Lewis acid , with the leaving diphosphate group itself playing the role of a general base and activating the incoming acceptor hydroxyl group for nucleophilic attack (Lairson and Withers, 2004). Enzymatic characterisation of this class of enzyme has been a long and daunting task. Conclusive identification of a catalytic nucleophile and the observation of a covalent intermediate have yet to be reported for any retaining transferases. Despite numerous attempts on a variety of retaining GTs, the sole evidence for an intermediate formation was seen in the structure of a mutant α-1,4-Galactosyltransferase (Lairson *et al.*, 2004). However, the result observed from this structure is quite confusing as the intermediate formation involved an amino acid located rather far, about 9 Å away from the reaction centre of the enzyme. In the absence of any conclusive evidence to support the double displacement theory, an alternative mechanism for these retaining GTs has been put forth. This new proposed mechanism suggests that a direct attack by the acceptor takes place, as previously proposed for glycogen phosphorylase (Klein *et al.*, 1986). This new mechanism termed S_Ni mechanism was first proposed by Persson *et al.* (2001) for bacterial LgtC. Retaining GTs which possess a GT-B fold that have distant structural similarities with glycogen phosphorylases, have been proposed to adopt the same mechanism (Gibson *et al.*, 2002).

A common feature in all retaining GTs is that the nucleotide sugar adopts a distorted folded conformation in which the pyranose ring of the sugar is folded below the plane of the diphosphate moiety of the nucleotide. This conformation is thought to facilitate hydrolysis of the sugar at the C1 position and promote the transfer of the sugar to the acceptor molecule. This can be achieved by several means: the C1 carbon is spatially accessible for the reaction, the folded orientation causes stress and weakens the anomeric bond at the C1 position hence making it more vulnerable to nucleophilic attack and a hydrogen bond can be established between O2 of the sugar and phosphate, lowering the energy barrier (Breton *et al.*, 2006).

2.1.4 Family 6 Glycosyltransferases: α-1,3-Galactosyltransferase and Forssman synthase

All the enzymes that are involved in the biosynthesis of the Gal(NAc)α1-3Gal(NAc)-containing antigens display strong amino acid sequence similarities (Figure 2.6), suggesting a common goal and all belong to the same family, GT6, according to the classification in the CAZy database (Heissigerova *et al.*, 2003). Enzymes belonging to this α3-Gal(NAc)T family all share the same features. They all utilise a UDP-nucleotide sugar as donor, retain the configuration of the Gal (or GalNAc) transferred, and their catalytic activity is strictly dependent upon the presence of a divalent cation (generally manganese) (Heissigerova *et al.*, 2003). These enzymes differ by their specificity: α-Galactosyltransferase (α3GT), blood group B transferase and iGb_3-S only use UDP-Gal as donor, while blood group A transferase and Forssman synthase (FS) use UDP-GalNAc. In addition to this the mouse AB glycosyltransferase can transfer sugar from both donors (Heissigerova *et al.*, 2003; Yamamoto *et al.*, 2001). The sequence and structural basis for Gal versus GalNAc specificity has been well reported for the blood group A and B transferases, which differ by only 4 amino acids (Persson *et al.*, 2007; Yamamoto and McNeill, 1996). However, there is limited information on the basis for acceptor substrate specificity in which α3GT and iGb_3-S utilise N-acetyllactosamine and lactose respectively while blood group enzymes require substitution by a L-fucose on position 2 of the acceptor galactose, whereas FS requires a N-acetyl group at the same position (Table 2.1).

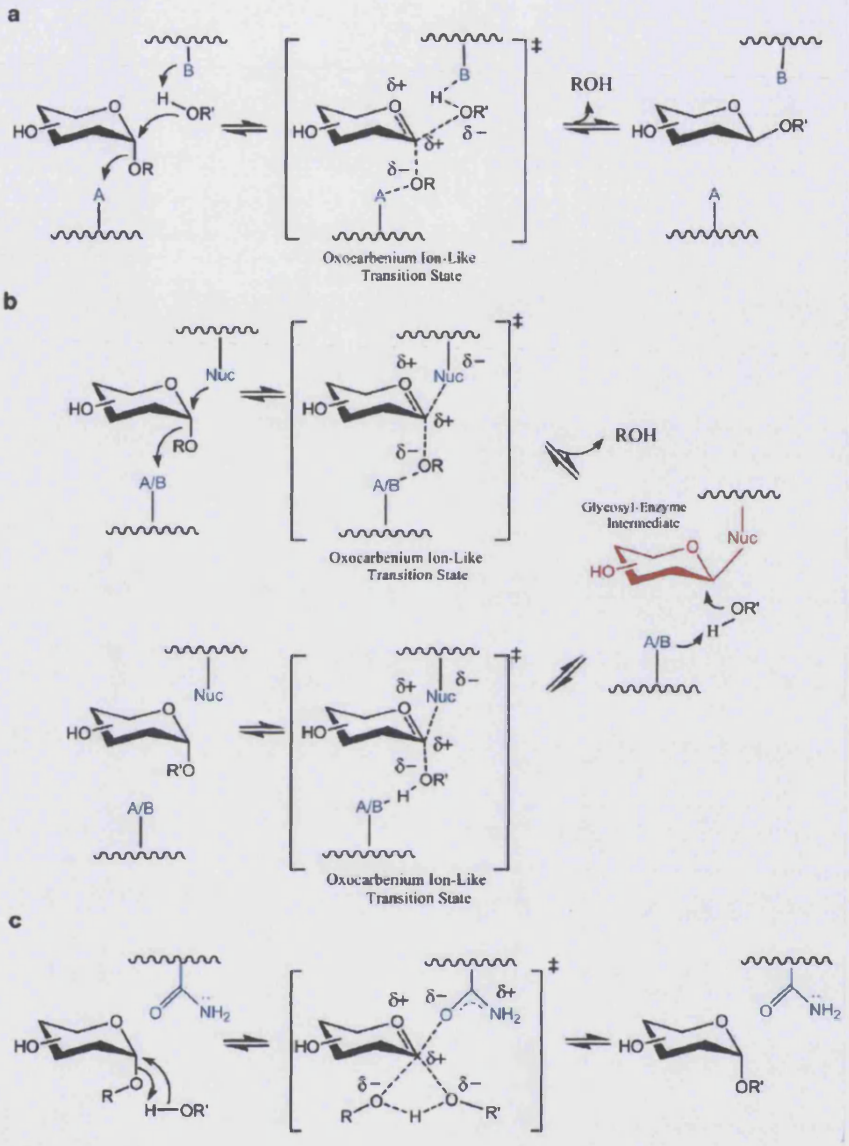


Figure 2.5: Established and proposed mechanisms for glycosidases/transglycosidases and glycosyltransferases, respectively. For glycosidases, R = a carbohydrate derivative and R'OH = H_2O or phosphate (phosphorylases classified as glycosidases). For transglycosidases, R=a carbohydrate derivative and R'OH = another sugar. For glycosyltransferases, R = a nucleoside diphosphate (e.g. UDP, GDP), a lipid phosphate, or phosphate (phosphorylases classified as glycosyltransferases) and R'OH = an acceptor group (e.g. another sugar or a protein). (a) A direct displacement S_N2-like reaction results in inverted anomeric configuration via a single oxocarbenium ion-like transition state. (b) The double displacement mechanism proceeds via two oxocarbenium ion-like transition states with the intermediate formation of a discrete covalently bound glycosyl–enzyme species, resulting in overall retention of anomeric configuration. (c) S_Ni-like mechanism proposed for retaining glycosyltransferases and glycogen phosphorylase involving a direct frontside displacement resulting in retention of anomeric configuration. (Lairson and Withers, 2004).

Enzyme	Donor	Acceptor	Product
α3GT	UDP-Gal	β**Gal**1,4βGlcNAc-R	αGal1,3 βGal1,4βGlcNAc-R
GTB	UDP-Gal	αFuc1,2β**Gal**1,4βGlcNAc-R	αGal1,3 [αFuc1,2]βGal1,4βGlcNAc-R
iGb3-S	UDP-Gal	β**Gal**1,4βGlcβ-ceramide	αGal1,3 βGal1,4βGlcβ-ceramide
GTA	UDP-GalNAc	αFuc1,2β**Gal**1,4βGlcNAc-R	αGalNAc1,3[αFuc1,2]βGal1,4βGlcNAc-R
FS	UDP-GalNAc	β**GalNAc**1,3αGal-R	αGalNAc1,3 βGalNAc1,3αGal-R

Table 2.1: Donor substrates, acceptor substrates and reaction products of GT6 family enzymes. The monosaccharides that carry the acceptor hydroxyl are indicated in bold (Heissigerova *et al.*, 2003).

The crystal structures of 3 members of the family 6 glycosyltransferases have been elucidated thus far: bovine α-1,3 galactosyltransferase and human A and B transferases. These structural determinations allow the understanding at the atomic level of the interactions of individual amino acid residues with donor and acceptor carbohydrate hence facilitating the understanding of the mechanism of these enzymes. This thesis will move on to discuss in more detail 2 members from this family: Bovine α-1,3-galactosyltransferase (α3GT) and canine Forssman synthase.

Figure 2.6: Amino acid sequence alignment of bovine α3GT and related enzymes of the GT6 family. Conserved amino acids have a red background and preserved ones a green background. Regions involved in ligands binding are boxed.

2.1.5 α-1,3-galactosyltransferase (α3GT)

Biological function

α-1,3-Galactosyltransferase (α3GT) is an enzyme that catalyzes the transfer of galactose from UDP-α-D-Galactose (donor substrate) to N-acetylglucosamine (acceptor substrate) to form an α-1,3 linkage with β-galactosyl groups in glycoconjugates, also termed as α-Gal epitopes. The enzyme is expressed in many mammalian species but is absent from humans, apes and old world monkeys. Most of the genes expressed in other mammals have counterparts in humans. The α3GT gene however is an exception. Only partial sequences, similar to α3GT gene of non-primate mammals have been detected in the human genome and are considered to correspond to pseudo genes, due to multiple deletions leading to premature translation stops (Lanteri *et al.*, 2002). Attempts to detect human α3GT transcripts have remained unsuccessful thus far (Galili, 1999). Complete loss of expression of this gene in old world primates, enables the production of significant amounts of natural antibodies towards α-Gal (Galili *et al.*, 1987). These antibodies are strongly active in the presence of complement. This is thought to be advantageous as it permits the production of antibodies to provide a barrier against infection of pathogenic microorganisms that carry the galactoconjugate epitope. However, this presents a barrier against xenotransplantation which is one of the major approaches to dealing with the limited supply of donor organs. Transfer of organs from organisms which produce α3GT to organisms that lack its activity will cause hyperacute rejection and in particular, prevent projects of transplantation of pig organs to humans (Kobayashi and Cooper, 1999). Studies by Kuwaki and co-workers showed that elimination of α3GT activity in baboons prevented hyperacute rejection and extended survival of pig hearts in baboons for 2-6 months (Kuwaki *et al.*, 2005). Another study by Yamada and co-workers further supported these results in which they reported that by using α3GT knockout donors and a tolerance induction approach, they had successfully achieved life-supporting pig-to-baboon renal xenograft survivals of up to 83 days (Kuwaki *et al.*, 2005). Intravenous administration of α-Gal epitopes as oligosaccharides have been shown to result in

neutralisation of anti-Gal and concomitantly in a temporary delay of hyperacute rejection of pig organs (Simon *et al.*, 1998). Recent studies also showed that α3GT-expressing allogeneic tumour cell vaccine can be used for treatment of cancer cells. This cancer vaccine comprised irradiated allogeneic cancer cells transfected to express α3GT with antitumour activity. Vaccination results in α3GT-mediated expression of foreign α-Gal carbohydrate residues on cell membrane glycoproteins and glycolipids of the vaccine cancer cell "xenograft"; these foreign α-Gal epitopes, not present on human cells, then induce a hyperacute rejection of the vaccine pancreatic cancer cell xenograft. The hyperacute rejection involves pre-existing human anti-Gal antibodies (naturally occur against gut flora) that bind the foreign α-Gal epitopes, resulting in the rapid activation of complement and antibody-dependent cell-mediated cytotoxicity (ADCC). The host immune system then attacks pancreatic cancer cells, existant prior to vaccination and with cell membrane epitopes unmodified by α-Gal, resulting in ADCC and tumor growth inhibition or reduction (Chen *et al.*, 2001; Deriy *et al.*, 2005; Galili, 2005; Manches *et al.*, 2005). The limited information currently available regarding the structural basis of molecular recognition and catalysis by this enzyme prevents a clear understanding of the molecular basis of its biological functions in cellular interactions and regulation. Hence, high resolution structural data are needed for better understanding of the mechanism of this enzyme in order to facilitate the design of inhibitors and also for the enzymatic synthesis of glycoconjugates for therapeutic use in diseases.

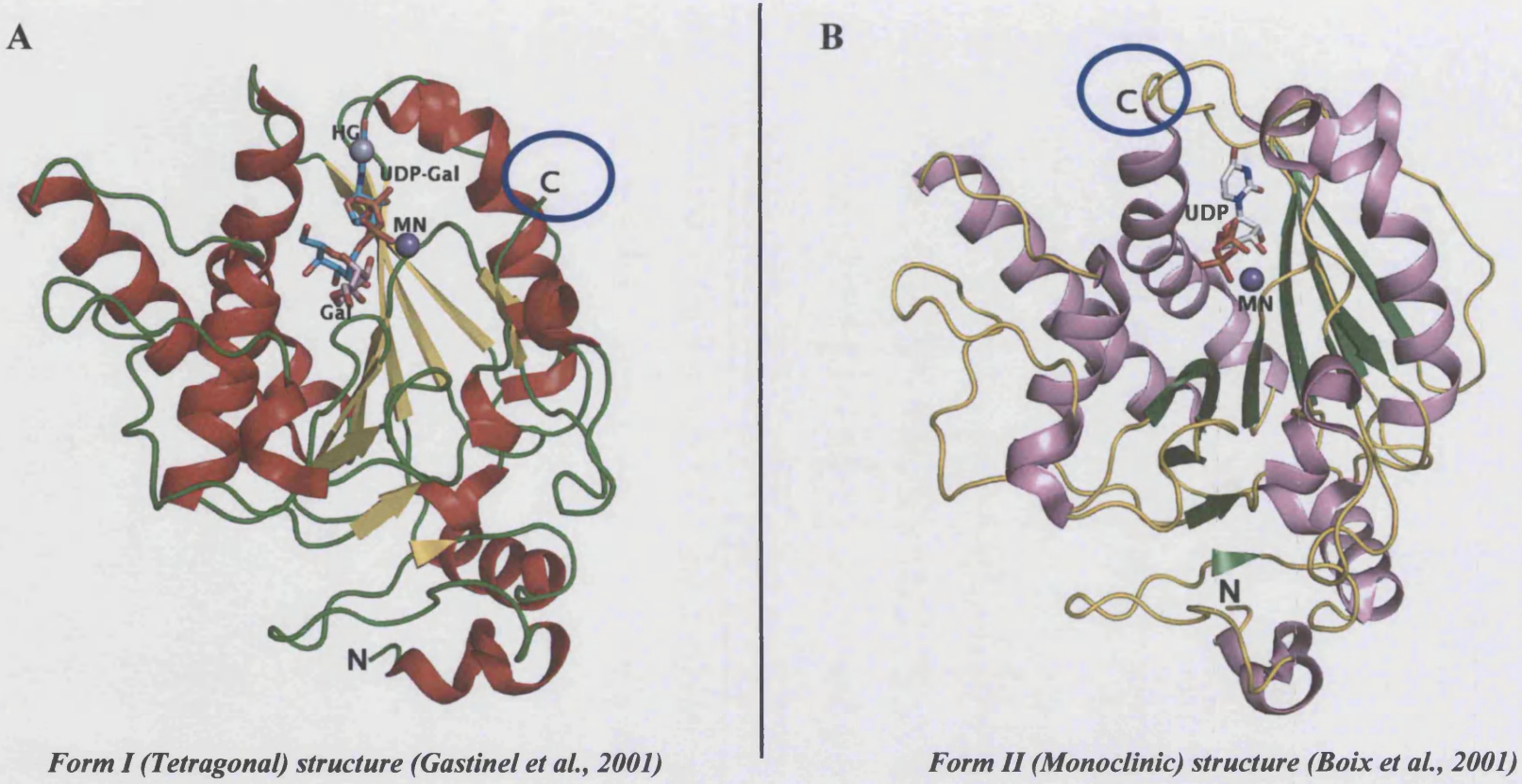
Structural studies of α3GT

The crystal structure of the catalytic domain (residues 80-368) of α3GT was first published by Gastinel and co-workers in the tetragonal form (form I), $P4_12_12$ space group at 2.3- and 2.5 Å resolution in UMP- and Hg-UDP-Galactose-bound forms, respectively (Gastinel *et al.*, 2001). The overall structure of the enzyme consists of 10 β-strands, 6 α-helices and 6 3_{10} helices. The folding of the protein is that of the typical α/β protein often observed in GTs with a central core of twisted β-sheet of 8 β-strands surrounded by 4 long α-helices. The central β-sheet can be divided into

two main regions. The first region runs from Val129 to Met224, defining the N-terminal subdomain and is made up of a β-sheet formed by 4 parallel β-strands and surrounded by two long α-helices. This N-terminal subdomain is the site for the binding of the nucleotide moiety of the nucleotide-sugar donor substrate. The second portion of the central β-sheet consists of 2 parallel β-strands flanked by 2 anti-parallel β-strands and with 2 long helices on one side. A second small β-sheet running almost parallel to the central β-sheet consists of 2 short antiparallel β-strands (Gastinel *et al.*, 2001). These structures identified binding sites for UDP and the manganese ion co-factor. A region of electron density in the Hg-UDP-Galactose-bound structure was interpreted as a β-galactosyl moiety covalently attached to residue Glu317 suggesting that a covalent catalytic mechanism takes place in the enzyme. However no direct evidence was obtained for a glycosyl-enzyme covalent bond, and the limited resolution of the structure raises questions about this interpretation (Boix *et al.*, 2001). This structure also exhibited disorder in the C-terminal region (residues 358-368), with this region having high B-factors of about 90 $Å^2$, suggesting that there is more than one conformation (Figure 2.7A).

In order to answer some of the questions that arose from the structure by Gastinel and co-workers, Boix and co-workers reported the structure of α3GT in complex with donor and acceptor substrates at 1.53 Å in the monoclinic space group $P2_1$ (form II) (Boix *et al.*, 2002; Boix *et al.*, 2001). This structure exhibited an ordered C-terminus and gave further insight into the catalytic mechanism of α3GT by revealing that there is a significant conformational change in the catalytically important C terminus during substrate binding (Figure 2.7B). The overall structure is almost identical to the form I structure reported by Gastinel and co-workers, however, the C-terminal region comprising residues 358-368, which is slightly disordered in the form I structure has undergone a large conformational change. This stretch of molecule is also highly ordered in form II as evidenced by the much lower *B*-factors reflecting the high resolution and much lower mobility for all residues. In the form II structure the C-terminus forms a lid over the active site of the molecule minimising the active site accessibility. In this structure the C-terminus forms a short

β-strand, and residues 361-364 adopt a definite α-helical structure and are well defined. Some of the residues in this form II structure, especially Arg365 make direct interactions with the UDP molecule in the active site (Figure 2.8).



Form I (Tetragonal) structure (Gastinel et al., 2001)

Form II (Monoclinic) structure (Boix et al., 2001)

Figure 2.7: Comparison between the Form I and Form II structures of α3GT. Root-mean –square-deviation of 0.17 Å for the form I and form II structures. The marked difference is in the conformation of the C-terminal region (circled in blue). In the form I structure (A), the C-terminal region is disordered and is shown to face away from the active site cleft whereas in the for II structure (B) the C-terminal region is ordered and is observed to fold over the active site cleft to form a pocket. Also shown in the form I structure is the bound mercury ion shown in a grey sphere and the galactose (shown in pink sticks) which is thought to make covalent interactions with the putative catalytic nucleophile Glu317 (Boix *et al.*, 2001;Gastinel *et al.*, 2001).

Mutational and biophysical studies showed that the C terminal region of α3GT is important for catalysis, possibly reflecting its role in a donor-substrate induced conformational change. Arg365 is a key residue on the C-terminus that interacts directly with the oxygen atom of the α-Phosphate of UDP (Figure 2.8). Mutation of this residue to lysine which is expected to perturb the conformational change of the C-terminus, showed that the enthalpy of binding for UDP is greatly reduced, in keeping with the role of Arg365 in UDP binding, while k_{cat} for UDP-Galactose hydrolysis and galactosyl transfer to lactose are reduced 12- and 40-fold, respectively (Zhang *et al.*, 2003). These results suggest that the conformational change in the C-terminal region of α3GT has a role in the rate-limiting step for these reactions.

Subsequent structural and calorimetric binding studies suggest an obligatory ordered binding of donor and acceptor substrates, linked to a donor substrate-induced conformational change, and the direct participation of UDP in acceptor binding (Boix *et al.*, 2002). The crystal structures and molecular modeling suggest that binding of a distorted conformation of UDP-Gal may be important in the catalytic mechanism of α3GT.

Several other mutational studies have been carried out in order to study the catalytic mechanism of α3GT (Boix *et al.*, 2002; Zhang *et al.*, 2004; Zhang *et al.*, 2003). One of the hypothesis that was put forth with the elucidation of the form I structure of α3GT is that the mechanism of the enzyme involves the formation of glycosyl-enzyme intermediate characteristic of a double displacement mechanism. However due to the low resolution of the structure as well as the lack of any direct evidence supporting this, the hypothesis has been cast in doubt. Subsequent mutational and structural studies that were carried out were designed in order to better understand that catalytic mechanism of the enzyme and to determine the exact role of Glu317 which was previously purported to be the catalytic nucleophile in the double displacement mechanism. In a study by Zhang *et. al* (2003) mutation of Glu317 to Gln was observed to weaken lactose binding and reduces the k_{cat} for galactosyltransfer to lactose and water by 2400-fold and 120-fold, respectively. The structure is not perturbed by this substitution, but the orientation of the bound lactose

molecule is changed. However, the magnitude of these changes does not support a previous proposal that Glu317 is the catalytic nucleophile in a double displacement mechanism and suggests it acts in acceptor substrate binding and in stabilizing a cationic transition state for cleavage of the bond between UDP and C1 of the galactose (Zhang *et al.*, 2003). Several other mutations were carried out on tryptophan residues which line the acceptor binding region of α3GT and this study showed that the mutations reduced both hydrolase and transferase activity and reduces affinity for the acceptor substrate (Zhang *et al.*, 2004). Crystal structures of some of these mutants did not show any changes in the overall structure of the enzyme however, localised changes that occur in the interactions of the mutated residues with substrates affect the catalytic activity and substrate binding of α3GT. Both mutational studies highlighted the importance of hydrogen bonds and hydrophobic interactions between enzyme and substrates in α3GT. These interactions play a pivotal role in arranging substrates in appropriate conformations and orientation for efficient catalysis.

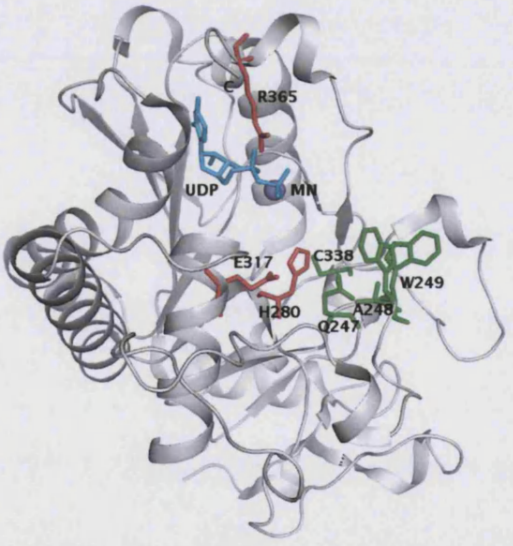


Figure 2.8: The figure shows the mutation sites of the mutants initially studied. Subsequent mutants that were produced, structurally studied and discussed in this thesis were based on the complex structures of these initial mutants. The UDP is shown in cyan and the manganese ion as a purple sphere. The individual residues are represented in sticks with the green residues representing the mutation sites for the 5-site mutant: Trp249, Ala248, Gln247, Cys338 and also includes a mutation of His280. The reason for studying the 5-site mutant is to determine the importance of the conserved cys338 and also to study the affect that mutation has on donor substrate and acceptor substrate binding.

2.2 Aim of studies

The main goal of this study is to carry out structural studies of C-terminal and active site mutants of α3GT in order to understand the significance of the conformational flexibility of some regions in the enzyme especially the C-terminal region. The aims are to determine the actual role of the C-terminus in the catalytic activity of the enzyme as well as determining the mechanism of α3GT. The other aim of this study is to determine the crystal structure of α3GT in complex with the inhibitor, U66. This study also extends to determine the crystal structure of Forssman synthase enzyme which is another enzyme from the glycosyltransferase 6 family, in complex with its substrates.

CHAPTER 3

Structural Studies on Bovine α-1,3 Galactosyltransferase (α3GT)

3.1 Methodology

3.1.1 α3GT sample preparation

All α3GT protein samples (wild type and mutants) and the U66 inhibitor were kindly provided to us by our collaborator Professor Keith Brew from Florida Atlantic University, USA. The purification process for α3GT has been published (Zhang *et al.*, 2001). Both proteins are unstable and at extended period of storage would precipitate out of solution, hence, the proteins are stored in a stable buffer containing 50% glycerol and 20 mM MES-NaOH, pH 6.5. Due to high concentration of glycerol in the buffer, the protein samples need to undergo a preparation process which involves buffer exchange and concentration prior to crystallisation. The protein sample is buffer exchanged into a buffer containing significantly less glycerol but still high enough to keep the protein stable in solution for crystallisation. The protein preparation described here is based on a previous method carried out by Boix and co-workers for the crystallisation of wild type α3GT (unpublished data). The protocol described is for the preparation of 250 μl of protein at concentration of 2-5 mg/ml. The buffer measurements are altered accordingly according to the volume of the starting protein material being used. All protein samples both for the wild type and mutants of α3GT were prepared using this protocol unless otherwise stated:

Buffer exchange and concentration

Concentrators:

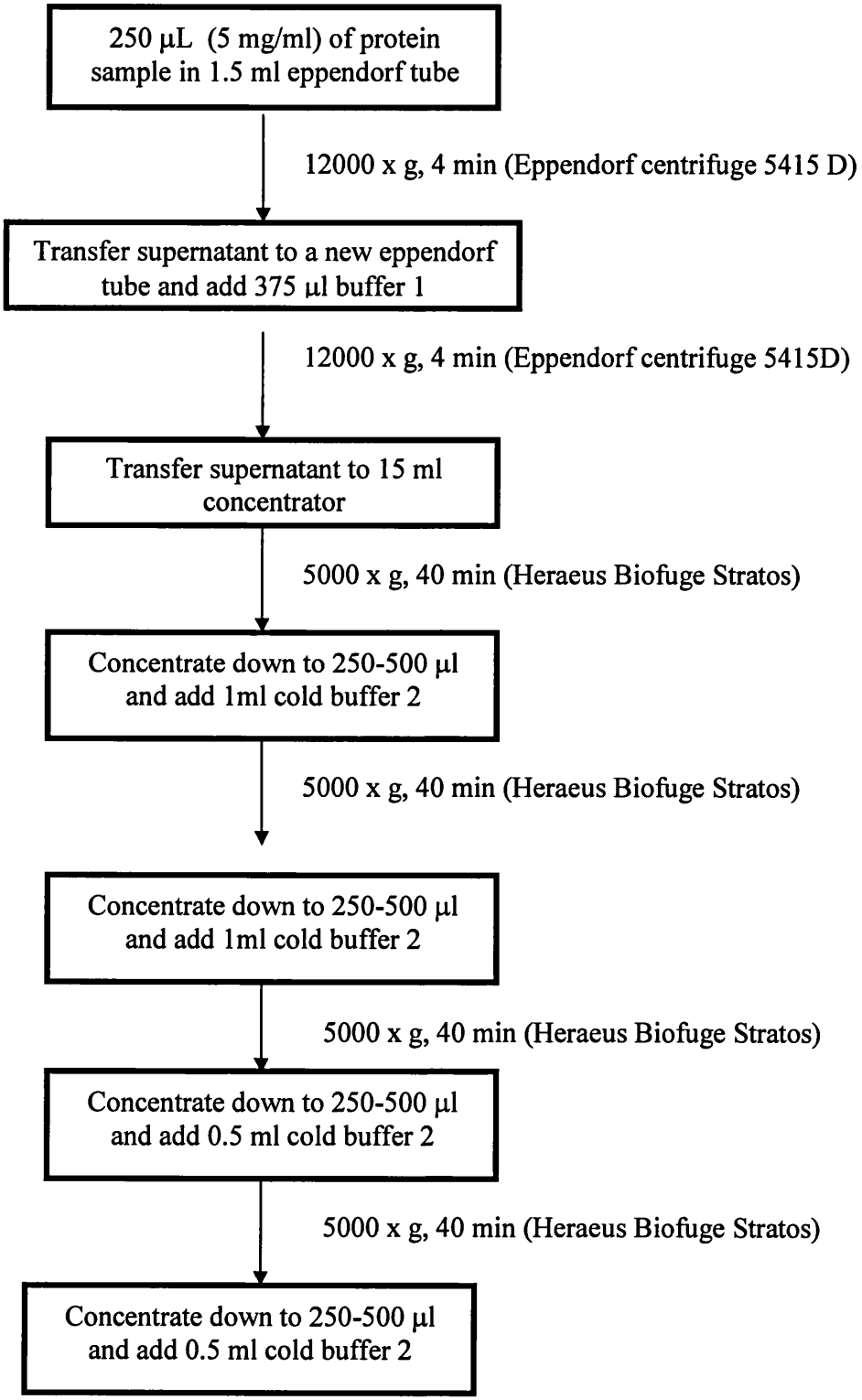
15 ml volume concentrators: Millipore UFV2BGC10 Ultrafree-15 centrifugal filter device; Biomax 10K (Ref. Z36, 463-0 from SIGMA)

Eppendorf 0.5 ml volume concentrators: IVSS vivaspin 500 centrifugal concentrator filter device, 10K (Ref. Z614025-25E, VS0101 from SIGMA)

Buffer preparation:

Buffer 1: 375 μl 20mM MES, pH6

Buffer 2 (for 10 ml): 0.2 ml MES 1M, pH6, 1 ml Glycerol 100%, 8.8 ml Analar H_2O (keep on ice)



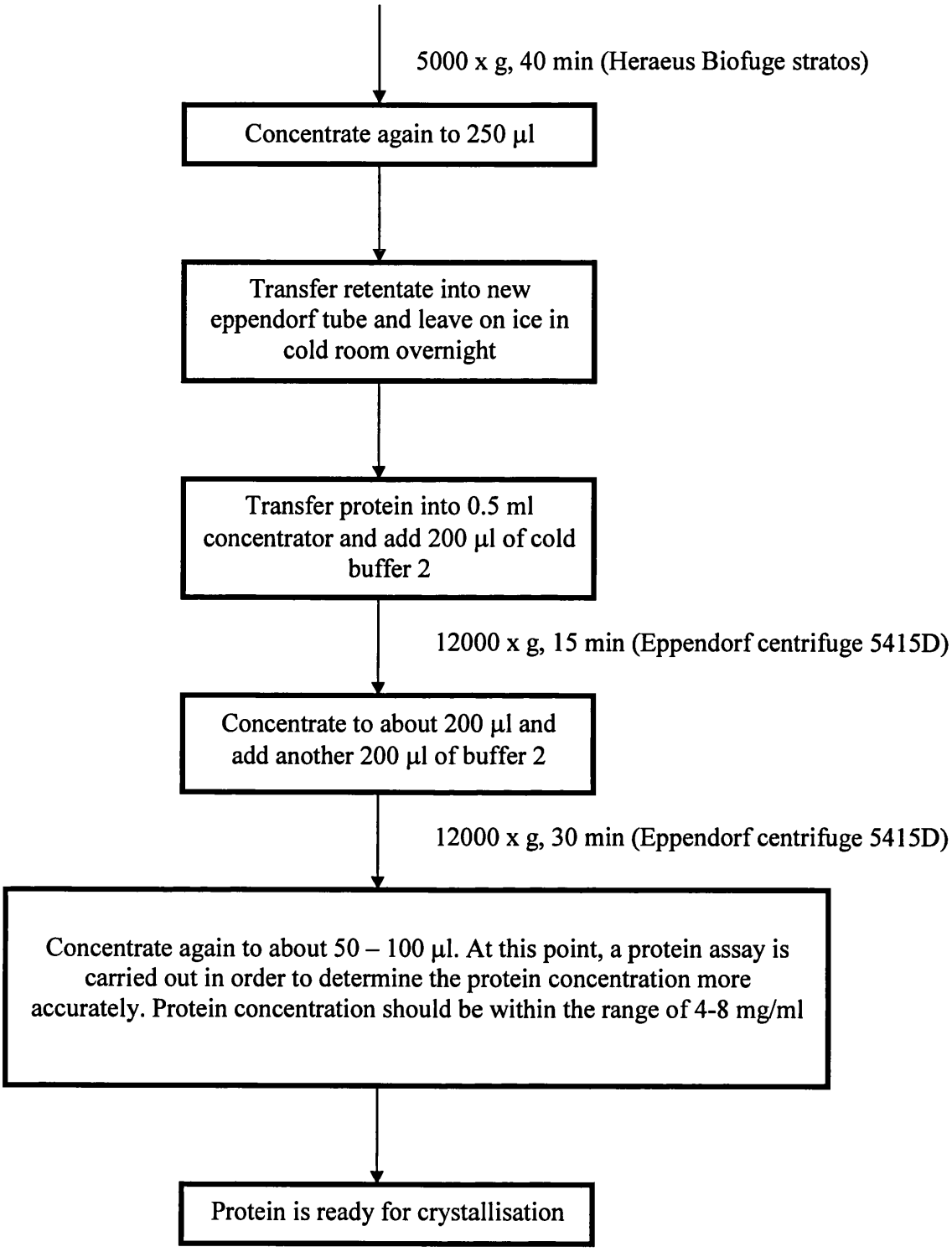


Figure 3.0: Flow chart describing the steps involved in the buffer exchange process of α3GT.

Protein estimation with Bradford Assay:

Protein estimation was carried out using a method based on Bradford (1976) (Bradford, 1976). The final volume of each assay is 1 ml. Bovine serum albumin (BSA) was used as a standard and a series of concentration in the range of 0.01 to 0.1 μg/μl was prepared to a final volume of 50 μl with analar water. Aliquots of 0.5, 1 and 2 μl of the α3GT sample were also used and diluted to 50 μl. 950 μl of Bradford reagent was added to the BSA standards and α3GT samples. Both standards and samples were left at room temperature for 10 minutes before the absorbance was recorded spectrophotometerically (Pharmacia Biotech, Ultrospec 2000 UV/Vis spectrophotometer) at 562 nm. A standard graph of absorbance versus the concentration of protein was plotted (Figure 3.1). The amount of protein present in the α3GT samples can then be estimated from the graph.

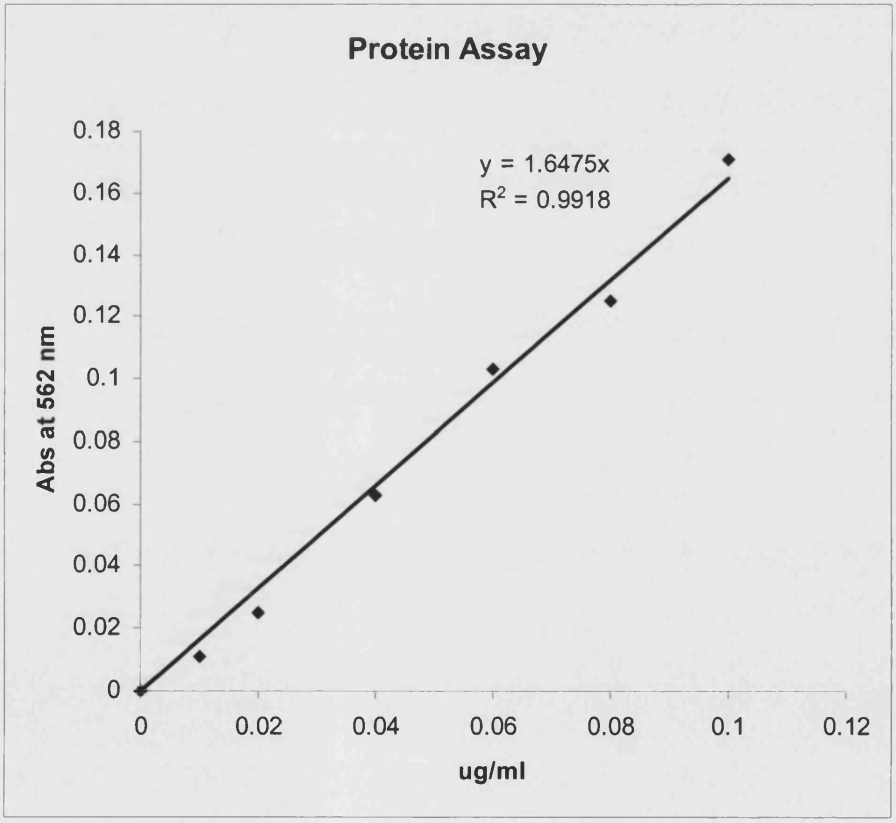


Figure 3.1: Protein assay calibration curve using known concentration of BSA

3.1.2 Crystallisation of α3GT mutants

Crystals were grown at 16°C by the vapour diffusion hanging drop method. 1μl of the protein in 20mM MES-NaOH buffer, pH 6 and 10% glycerol, 10mM donor substrates (UDP or UDP-Gal Sigma®) and 10mM manganese chloride was mixed with 1μl of reservoir solution. Crystallisation was also carried out in the presence of acceptor substrate N-acetyllactosamine (LacNAc) (Sigma®).

- **Streak seeding:**

Crystallisation drops were set up as described above. The drops were then pre-equilibrated at 16°C for about 1 hour. A fine fibre (Cat whisker) was used to pick up microscopic crystal seeds from an existing crystal. The fibre was then used to seed the pre-equilibrated drop by stroking the fibre in a straight line through the drop.

- **Crystallisation conditions and screens**

Known α3GT crystallisation conditions:

Monoclinic form(Boix *et al.*, 2002; Boix *et al.*, 2001): 5-10% PEG 6000 (Sigma®), 0.1M Tris-HCl, pH 8.0 (form II)

Tetragonal form (Zhang *et al.*, 2003): 1.4-2.0M $(NH_4)_2$ SO_4, Tris-HCl, pH 8.0 (form I)

Crystallisation screens:

Structure Screen I and II from Molecular Dimensions Limited, UK and PEG/Ion screen from Hampton Research Corp., USA.

Self designed crystallisation screen:

Crystallisation screen designed based on the monoclinic (form II) crystallisation condition (Boix *et al.*, 2002; Boix *et al.*, 2001):

Several changes were carried out according to the results obtained after setting up crystallisation based on the known α3GT conditions and also the conditions from the standard crystallisation screens.

Main variables that were changed:

- Drop size (protein to reservoir ratio of: 1:1, 2:1, 1:2 etc.)
- Type of PEG (PEG 3350, 4000, 6000, 8000)
- Concentration of PEG (concentration range of 1% to 35%)
- pH of buffer (pH range 7.0-9.0)
- Additives, e.g: Salt (i.e: Sodium citrate, Sodium acetate etc.), organic solvents (i.e: MPD, Jeffamine, Ethanol etc.)

3.1.3 Results of Crystallisation experiments

Crystallisation experiments often go through two distinguishable phases: a screening step, which is frequently followed by optimisation. Screening establishes which conditions produce promising crystals – or at least, some leads – and optimisation refers to the fine tuning of those initial conditions to obtain well diffracting crystals which are suitable for data collection.

Wild type α3GT

One of our objectives is to co-crystallise wild type α3GT with the inhibitor U66, however, initial attempts to co-crystallise wild type α3GT with the inhibitor U66 were unsuccessful. On addition of U66, α3GT precipitates and it is concluded that the best way to overcome this problem was to grow α3GT crystals without substrates or co-factor (apo form) and then carry out soaking experiments with U66 (Results for experiments with U66 are discussed in section 3.4).

α3GT mutant complexes

Apart from carrying out co-crystallisation attempts of wild type α3GT with the inhibitor U66, another aim of this research was to obtain the crystal structures of mutants of α3GT in complex with donor substrates and donor substrate analogues for the purpose of understanding the mode of binding of these substrates hence

shedding light into the catalytic mechanism of α3GT. Previously, crystallisation of α3GT mutants primarily the Glu317Gln and the active site tryptophan mutants were carried out in the presence of UDP (Zhang *et al.*, 2004; Zhang *et al.*, 2003) hence often resulting in the structures of the mutants in complex with UDP which is a dead end inhibitor of the enzyme. As co-crystals of the α3GT mutants in complex with UDP-Gal did not grow in the original crystallisation condition reported by Boix and co-workers (Boix *et al.*, 2001), a new crystallisation condition which promotes the formation of co-crystals of the mutants in complex with the donor substrate UDP-Gal was sought.

Attempts were carried out to co-crystallise the α3GT mutants with UDP-Gal and manganese ions. Preliminary crystallisation screens with commercial screens from Molecular Dimensions and Hampton research resulted in the identification of a few new conditions which produced crystals. These conditions in general produced bypyramidal looking crystals which are usually associated with the tetragonal crystal form. Unfortunately, these bipyramidal crystals did not diffract very well and did not show any ligand binding in the active site. This result is in agreement with the tetragonal structure reported by Boix and co-workers (Zhang *et al.*, 2003) which suggested that the tetragonal crystal form of α3GT is the open inactive form of the protein. This result indicated that a different crystallisation condition should be pursued especially a condition which favours crystal formation in the active form or at the very least, a condition which produced crystals in a space group other than tetragonal. It was observed that crystallisation conditions that utilised high salt concentration will often produce α3GT crystals in the tetragonal space group.

At this point, conventional commercial screens were abandoned and a new screen was formulated based on crystallisation condition for the monoclinic crystal form reported by Boix and co-workers (Boix *et al.*, 2002; Boix *et al.*, 2001). This condition utilised PEG 6000 as the precipitant and Tris as buffer. A crystallisation screen with different types of PEG and incorporation of additives listed in table 3.0 was formulated.

Additive	Concentration	Results
Salts (chloride, sulphate, acetate and citrate)	0.05M-0.3M	Several conditions containing acetate and citrate salt produce a shower of micro-crystals
Ethanol	2-15%	Precipitates
Methanol	2-15%	Precipitates
hexafluoropropanol	1-5%	Precipitate
2-propanol	2-15%	Precipitates
DMSO	2-15%	Heavy precipitate
DTT	2-15%	Heavy precipitate
Dioxane	2-15%	Light precipitate
MPD	2-30%	Crystal plates of reasonable size
1,6-hexanediol	2-15%	Precipitates
1,2,3-heptane diol	2-15%	Precipitates
xylitol	5-30%	Precipitate
Ethylene glycol	5-20%	Precipitates
1,2,3-butanediol	2-15%	Precipitate
Jeffamine	2-15%	Light precipitate and small crystals
PEG 200-600	2-15%	Precipitates

Table 3.0: The list of additives used to screen for a co-crystallisation condition for the α3GT mutants. The concentrations shown are the final concentrations in the crystallisation well containing 5-15% PEG 6000 and 0.1M Tris-HCl, pH 8, hence the concentration of the additives in a 2 micro-liter drop is half of the concentration shown.

The addition of additives especially salt was vital as salt assists the solubility of the protein as previous attempts had shown that the α3GT mutants precipitated immediately when crystallised in the presence of just PEG and Tris buffer. The new screen gave more promising results as co-crystals of the mutants with UDP-Gal were obtained in conditions containing 5-10% PEG 6000, Tris pH 8 and 5-20% MPD. Interestingly for most of the mutants, MPD seemed to be a common and perhaps an important factor for crystal growth as the original condition without MPD did not produce any crystals, in fact co-crystallisation attempts of the mutants with UDP-Gal in the original condition would usually result in precipitation. MPD seemed to be very important for crystal contacts as a result of adding which all the crystallisation conditions, despite varying the concentration of PEG 6000 and even the pH of the Tris buffer, will produce diffraction quality crystals. In general subsequent crystallisation of α3GT mutants unless otherwise stated were carried out using this condition with variations in the concentration of PEG 6000 and MPD.

3.2 Conformational dynamics of α-1,3 Galactosyltransferase (α3GT)

3.2.1 Introduction

To understand the structural basis of catalysis and substrate specificity in α3GT, our goal is to elucidate the interactions between the substrates, metal cofactors and products with the enzyme as well as the effects of mutations of key residues in the active site. Initial crystallographic studies of a tetragonal crystal form (form I) of a complex of α3GT with a Hg-derivative of UDP-Gal suggested the presence of a galactosyl moiety, covalently bound to Glu317 (Gastinel *et al.*, 2001), consistent with the view that Glu317 might act as the catalytic nucleophile in a double displacement mechanism, the expected mechanism for a retaining glycosyltransferase (Gastinel *et al.*, 2001; Unligil *et al.*, 2000; Zhang *et al.*, 2004; Zhang *et al.*, 2003; Zhang *et al.*, 2001). However, this observation was puzzling since the structure also contained a second galactose that was part of a non-covalently bound Hg-UDP-Gal, and the electron density for the galactose adjacent to Glu317 was weak. Subsequent kinetic, structural and mutational studies do not support a double-displacement mechanism for α3GT.

The relatively lower resolution tetragonal crystal form (form I), had a disordered C-terminal region (Gastinel *et al.*, 2001) and a higher resolution monoclinic crystal form (form II), represented by the enzyme complex with UDP and manganese, was later solved in which the C-terminal region has a distinct and highly ordered conformation(Boix *et al.*, 2001). Since then we have reported form II structures for α3GT in complexes with manganese and UDP, manganese/UDP-Galactose (UDP-Gal), manganese/UDP-Glucose (UDP-Glc) and with both manganese/UDP and the acceptor substrates lactose (LAC) and *N*-acetyllactosamine (Lac NAc) (Boix *et al.*, 2002). In the complexes containing UDP-Gal and UDP-glc, the bond between the UDP and monosaccharide was cleaved, leaving a β-Galactose and α-Glucose trapped in the active site. Because of the relatively high level of UDP-Gal hydrolase activity of α3GT, we have been unable to determine the structure of an enzyme complex containing an intact UDP-Gal substrate. UDP-2F-Galactose (UDP-2F-Gal) (Figure 3.2) is a UDP-Gal analogue that is resistant to hydrolysis and monosaccharide transfer that has been shown

to be a reversible inhibitor of α3GT (K_i=245μM) and other GTs (Burkart *et al.*, 2000). Unlike the mercurial derivative of UDP-Gal, this derivative is closely similar in size to the donor substrate. We were unsuccessful in attempts to crystallize a complex of this inhibitor with wild-type α3GT, but have been able to crystallize and solve the structure of its complex with the Arg365Lys mutant. Previous studies showed that Arg365 in the C-terminus of α3GT interacts with the α-phosphate of UDP and that this mutation to Lys specifically perturbs catalytic activity but not substrate binding (Boix *et al.*, 2001; Zhang *et al.*, 2003). The structure of the complex with UDP-2F-Gal is the first structure of a complex for an enzyme in this family with a donor substrate analogue inhibitor. It provides high resolution information about the conformation of the donor substrate mimetic and enzyme as well as their interactions in what is presumably an appropriate model for the Michaelis complex with UDP-Gal. To extend the picture, we have also determined the structures of the apo-forms and UDP complexes of the wild-type enzyme and Arg365Lys mutant. Together, these show that progressive conformational changes occur in two regions of α3GT during the catalytic cycle; mutational and truncation studies suggest that conformational changes in the C-terminal region are important for catalysis. This is consistent with studies with other glycosyltransferases that suggest a role for loop conformations in modulating the active site environment during the catalytic cycle (Unligil and Rini, 2000).

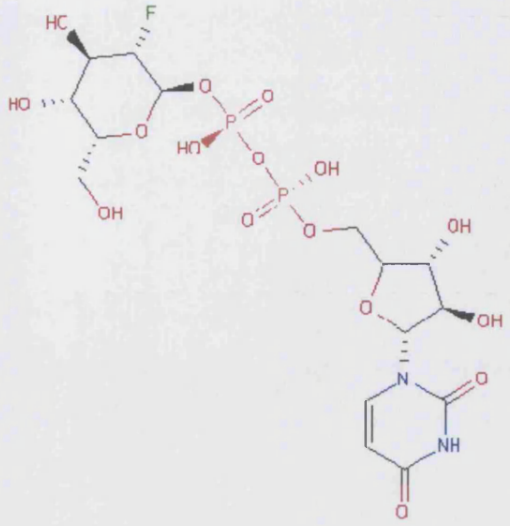


Figure 3.2: Chemical structure of UDP-2F-Gal. UDP-2F-Gal is more similar in size to the donor substrate, UDP-Gal with the difference in substitution of a fluorine atom at the 2' position of the galactose moiety (shown in green). [Chemical structure of UDP-2F-Gal obtained from the PDB entry: 2JCF (Jamaluddin *et al.*, 2007)]

3.2.2 Methods

Crystallisation

Co-crystals of α3GT mutants, Arg365Lys mutant and the C-terminal truncated mutant, CΔ3 in the presence of substrates and substrate analogues were grown in conditions containing 5-10% PEG 6000, 0.1M Tris-HCl, pH 8 and 5-20% MPD as discussed in section 3.1 while crystals of the apo form of wild type α3GT and the Arg365Lys mutant were grown in conditions containing Tris-HCl buffer pH 8.0 and 1.4M ammonium sulphate and 1.4M sodium acetate and 0.1M sodium cacodylate, pH 6.5 respectively. Crystallisation techniques for the apo form of the enzyme are discussed in more detailed in subchapter 3.4.

Data collection, Processing and Structure refinement

Datasets for the apo-form of the protein and a high resolution dataset for the Arg365Lys mutant to 1.8 Å were collected at the Synchrotron Radiation Source (SRS), Daresbury, UK (station PX 9.6 and PX 14.2 respectively, using an Area Detector Systems Corp. detector system). Diffraction data for the Arg365Lys mutant in a complex with UDP-2F-Gal to 2.4 Å resolution were collected at the SRS, Daresbury (station 14.1, with an MAR 345 image plate system). The dataset for the CΔ3 mutant in a complex with UDP-Gal was collected to 2.0 Å resolution at DESY, EMBL outstation, Hamburg, Germany (station X13, using a 30-cm MAR research image plate system). Raw data images for the wild type apo structure, Arg365Lys apo structure, Arg365Lys in complex with UDP and Arg365Lys in complex with UDP-2F-Galactose were indexed and scaled using the DENZO and SCALEPACK modules of the HKL suite (Otwinowski and Minor, 1997) while processing of the CΔ3 mutant dataset was carried out using MOSFLM (Leslie, 2001) and SCALA (CCP4, 1994).

All five structures were determined by the molecular replacement method using the native form II structure (PDB code: 1K4V)(Boix *et al.*, 2001) as a search model with the program MOLREP (Vagin and Teplyakov, 1997). Crystallographic refinement was performed using the program package CNS (Brunger *et al.*, 1998) for the complex structures of the mutants. Several rounds of energy minimization, simulated annealing, B-factor refinement and model building using the program COOT(Emsley, 2004) were performed until the R_{free} could not be improved. Appropriate ligand(s) were inserted into the structure as the refinement progressed. Residues with poor side-chain density were modelled as alanines and regions with very poor main chain electron density were excluded from the refinement. Water molecules were gradually included into the model at positions corresponding to peaks in the F_o–F_c electron density map with heights greater than 3σ and at H-bond distance from appropriate atoms. Occupancy for the second manganese binding site in the Arg365Lys mutant in the complex with UDP was determined by setting the B-factor of the manganese to the Wilson B-value of the data which is 17 $Å^2$ and then carrying

out a series of short refinement runs using different occupancies for the manganese ion (0.2 to 1.0). Refinement was carried out using the program REFMAC (CCP4, 1994) for the 'apo' form of the wild type enzyme and the Arg365Lys mutant structures. Data collection and refinement statistics for all 5 structures are listed in Table 3.1.

Protein	WT (apo)	R365K (apo)	R365K-UDP	R365K-UDP-2F-Galactose	CΔ3
Substrates used in crystallisation	none	none	10mM UDP 10mM $MnCl_2$	10mM UDP-2F-Gal 10mM GlcNAc 10mM $MnCl_2$	10mM UDP-Gal 10mM N-acetyllactosamine 10mM $MnCl_2$
Substrates observed in the crystal structure	none	none	2 manganese 1 UDP	2 manganese (one/monomer) 2 UDP-2F-Gal (UP1) (one/monomer)	1 manganese (one/monomer) 1 UDP (one/monomer)
C-terminus	Disordered	Disordered	Partially ordered	Disordered in molecule A Partially ordered in molecule B	Fully ordered (residues 358-365)
Space group	$P4_12_12$	$P2_12_12_1$	C2	P1	C2
Mols/A.S.U	1	2	1	2	1
Crystal parameters					
a (Å)	93.57	95.93	122.82	44.74	121.76
b (Å)	93.57	103.26	68.25	64.76	68.30
c (Å)	108.42	121.49	44.73	66.47	43.96
α (°)	90.0	90.0	90.0	83.3	90.0
β (°)	90.0	90.0	90.2	83.9	90.1
γ (°)	90.0	90.0	90.0	70.1	90.0
Resolution (Å)	50.00-2.56	50.0-3.3	50.0-1.8	50.0-2.4	50.0-2.01
R_{symm} (%)[a] (outermost shell)[b]	7.2 (68.3)	20.1 (61.9)	10 (21.0)	5.2 (12.8)	10.3 (34.1)
Completeness (%) (outermost shell)[b]	98.1 (84.8)	99.8 (99.9)	91.7 (93.6)	89.9 (67.2)	96.4 (80.8)
***I/σI* (outermost shell)[b]**	24.0 (2.1)	7.6 (2.4)	9 (5.1)	11.5 (4.4)	12.4 (3.2)
Total reflections	299696	220655	155949	111638	88715
Unique reflections	15988	18881	34457	27634	23156
R_{cryst}[c]/R_{free}[d] (%)	26.4/32.8	20.6/28.1	19.4/20.4	18.1/22.9	18.7/20.7
Ramachandran Plot % core/allowed	84.6/15.4	85.3/14.7	88.8/11.2	86.6/13.4	88.6/11.4
RMSD from ideal					
Bond angles (°)	1.73	1.63	1.39	1.30	1.42
Bond lengths (Å)	0.02	0.016	0.007	0.007	0.007
Water molecules	23	61	395	551	204
***B*-factor statistics (Å^2)**					
Protein[e]	58.9	37.0 (A), 35.0 (B)	15.4	30.6 (A), 31.3 (B)	16.5
Ligand in the active site[e]	-	-	13.5 (UDP)	47.7 (UP1-A), 33.8 (UP1-B)	16.8 (UDP), 33.5 (Tris)
Metal[e]	-	-	13.7 (MN1), 58.5 (MN2)	42 (MN1-A), 30.3 (MN1-B)	15.2 (MN1)

[a] $R_{symm}=\Sigma_h\Sigma_i[|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)]$, where Ii is the ith measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$.

[b] Outermost shell are 2.63 – 2.56 Å, 3.37 – 3.30 Å, 1.91 – 1.80 Å, 2.55 -2.40 Å and 2.07- 2.01 Å respectively .

[c] $R_{cryst}= \Sigma_h|F_o - F_c|/\Sigma hF_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h.

[d] R_{free} is equal to R_{cryst} for a randomly selected subset of reflections, not used in refinement.

[e] Temperature factors for individual molecules quoted

Table 3.1: Refinement statistics for the structures of the 'apo' form of wild type α3GT and Arg365Lys mutant, Complexes of the Arg365Lys mutant with UDP and 2 manganese ions, UDP-2F-gal and the structure of CΔ3 mutant in complex with UDP and Tris ion.

3.2.3 Results

Structures of apo- wild type α3GT and Arg365Lys mutant

The orthorhombic crystals of the apo-form of the Arg365Lys mutant with 2 molecules per asymmetric unit diffracted to 3.3Å and the structure has a higher average B-factor compared to the substrate bound structures (Table 3.1). In both molecules in the asymmetric unit, the C-terminus (residues 358-368) is disordered; in molecule A, there is also a high level of disorder in the Trp195 loop between residues 190 and 199 with the most significant differences from the UDP complex of wild-type α3GT (PDB code 1K4V) being in residues Trp195 and Gln196 (Figure 3.3). In molecule B, this region is more structured but residues 188 to 199 are arranged differently as a result of molecular contacts in the crystal. This suggests that, in the absence of ligands, this region is highly flexible.

Apo-wild type α3GT crystallised in the tetragonal space group with one molecule in the asymmetric unit, and diffracted to 2.56Å resolution. In general, the structure is closely similar to that of the apo form of the Arg365Lys mutant. The same two regions in the structure are disordered, the C-terminal 11 residues and Trp195 loop (residues 190-199). The disorder is reflected in a distinct break in electron density in the region corresponding to residues Trp195 and Gln196. The global disorder in the structure is reflected in the average B-factor which is close to 60 $Å^2$ (Table 3.1).

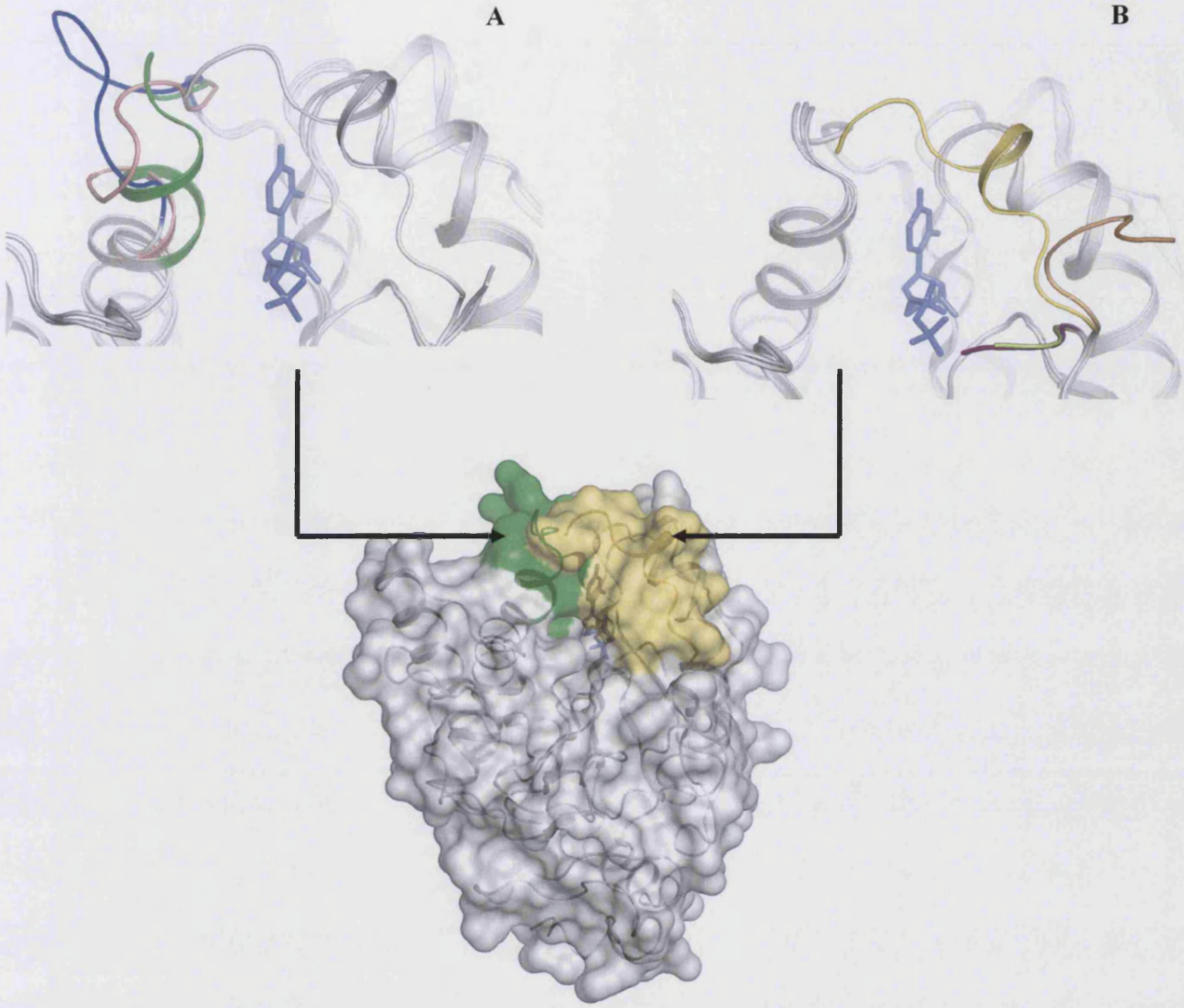
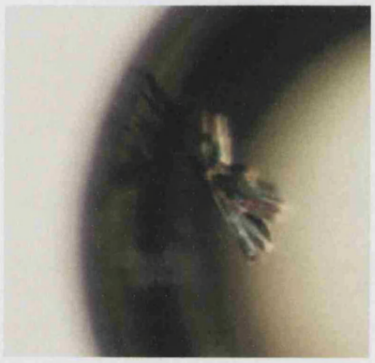


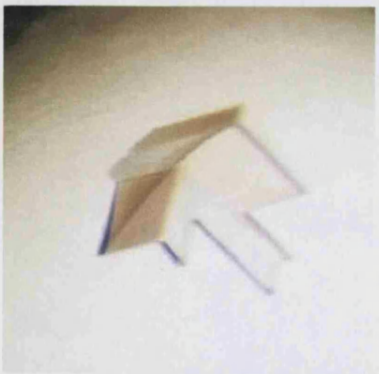
Figure 3.3: Surface representation of the wild type α3GT structure in complex with UDP (Boix *et al*; 2001, PDB code: 1K4V). The region highlighted in green represents the loop corresponding to residues Lys190-Ser199. The diagram in panel A shows the different conformation of this loop in the enzyme's apo form; apo wild type (blue) and apo Arg365Lys mutant (salmon) indicating that this region is highly flexible and is mostly disordered in the absence of donor ligand binding . The region highlighted in yellow shows the C-terminal region of the enzyme. Panel B shows the different conformation of the this region in different crystal structures of the mutants with different ligands bound in the active site; Arg365Lys mutant in complex with UDP-2F-Gal (orange), Arg365Lys mutant in complex with UDP and 2 manganese ions (light green) and C-terminal truncated mutant (CΔ3) in complex with UDP and Tris ion (purple).

UDP complex of Arg365Lys mutant

The crystals of the Arg365Lys mutant in complex with UDP were grown by mixing 1 µl of the protein (4-8 mg/ml) in 20mM MES-NaOH, pH 6.0, 10% glycerol, 10mM $MnCl_2$ and 10mM UDP with an equal volume of reservoir solution (Figure 3.4). Arg365Lys mutant crystals were also grown in conditions containing UDP-Gal using the same method as the co-crystallisation with UDP. These crystals diffracted to 1.8 Å and 1.77 Å respectively (Figure 3.5 & 3.6).



Crystals of Arg365Lys mutant crystallised in the presence of UDP and manganese. Crystallisation condition: 10% PEG 6000 and Tris-HCl pH 8. The crystal diffracted to 1.8 Å in space group C2. (The longest single needle is about 0.05 mm)



Clover leaf shaped crystal of Arg365Lys mutant crystallised in the presence of UDP-Gal and manganese. Crystallisation condition: 12% PEG 6000, 5% MPD, Tris pH 8. The crystal diffracted to 1.77 Å in space space group C2.
(Typical crystal dimensions: 0.1 X 0.2 X 0.2 mm)

Figure 3.4: Crystals of Arg365Lys mutant in complex with UDP and UDP-Gal in the C2 space group.

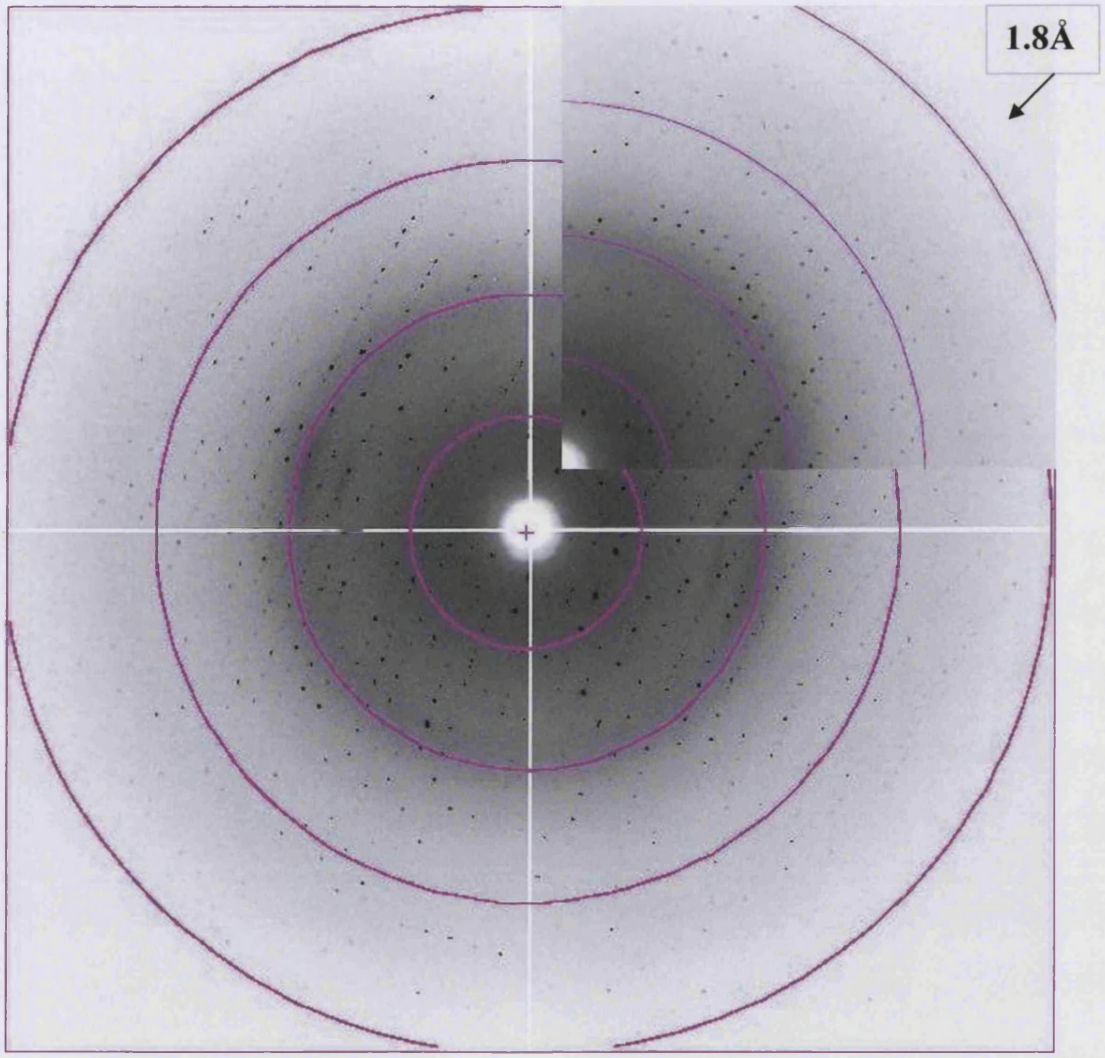


Figure 3.5: Diffraction image from Arg365Lys mutant crystal in complex with UDP and 2 manganese ions that diffracted to 1.8Å resolution. The inset represents a portion of the image zoomed in to show the high-resolution spots. The circles have been drawn to mark the resolution range. The data collection parameters used were: λ= 0.978 Å; $\Delta\varphi$= 1.0° and exposure = 15 seconds at SRS, Daresbury, UK (Station PX 14.2), collected using an ADSC-CCD detector. The rings mark the resolution limits of 7 Å, 3.6 Å, 2.4 Å and 1.8 Å proceeding outwards from the centre. The data was processed to 1.8 Å

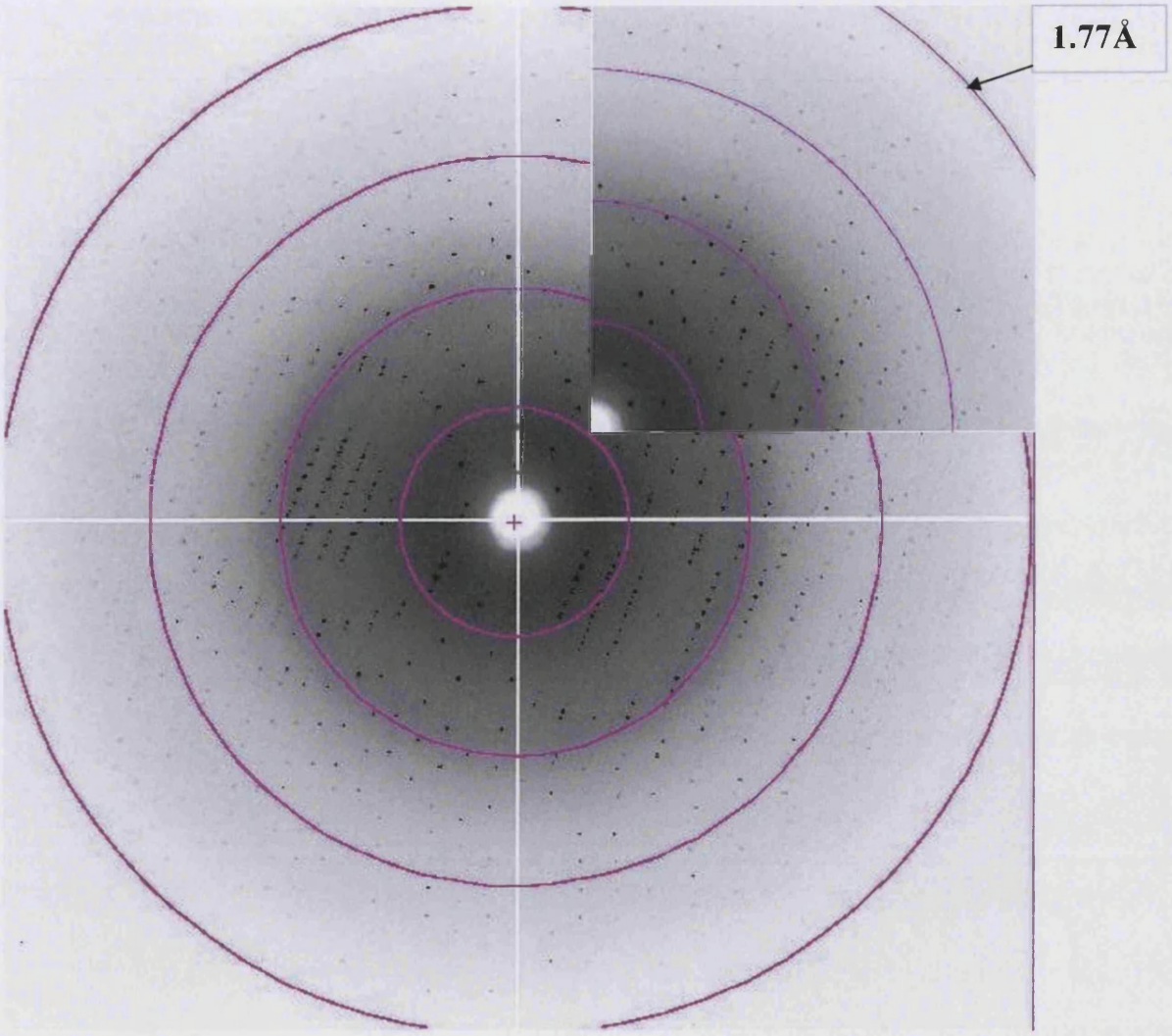


Figure 3.6: Diffraction image from Arg365Lys mutant crystal in complex with UDP-Gal and manganese that diffracted to 1.77Å resolution. The inset represents a portion of the image zoomed in to show the high-resolution spots. The circles have been drawn to mark the resolution range. The data collection parameters used were: λ= 0.978 Å; $\Delta\varphi$= 1.0° and exposure = 15 seconds at SRS, Daresbury, UK (Station PX 14.2), collected using an ADSC-CCD detector. The rings mark the resolution limits of 6 Å, 3 Å, 2 Å and 1.7 Å proceeding outwards from the centre. The data was processed to 1.77 Å

The crystals of these complexes of the Arg365Lys mutant are in the centred monoclinic space group with one molecule per asymmetric unit. The structure shows UDP and a manganese ion bound in the active site as in the wild-type α3GT structures previously described by Boix *et al* (2001). However, in the Arg365Lys•UDP complex structure, the C-terminal region is somewhat disordered and, initially, it appeared that there was no electron density for residues 358-368. However, with subsequent rounds of refinement and improved phases, density for the first three residues of the C-terminus became visible, enabling the modelling of Thr358, Lys359 and Glu360 into the structure. Inspection of this initial structure suggested that Glu360 interacts with the second phosphate of the UDP, which is chemically anomalous, but also showed the presence of a region of electron density that was not accounted for in the model. This density was subsequently identified as a second manganese ion, bound with about 90% occupancy at a location similar to that of the Tris ion in the structure of the UDP complex of the CΔ3 mutant (*vide infra*). Based on their proximity (<4 Å distance), the second manganese interacts with the side chains of Glu317, His280 and Gln247; the fourth ligand is a water molecule that bridges between the metal and the second phosphate of the UDP (Figure 3.7).

The substitution of Lys for Arg365 was not observed in the structure because of the disorder in residues 361-368; this may reflect the presence of the mutation, since Lys can form fewer H-bonds than Arg and also the presence of the second manganese that reduces interactions of Lys359 and Tyr361 with the second phosphate of the UDP.

We also crystallized the Arg365Lys mutant in the presence of UDP-Gal. The structure for this complex is identical to the structure of Arg365Lys in complex with UDP indicating that the UDP-Gal has undergone hydrolysis (unpublished results) in keeping with the significant UDP-Gal hydrolase activity of the Arg365Lys mutant (Zhang *et al.*, 2003; Zhang *et al.*, 2001). However, unlike the structure of the wild-type enzyme complex with UDP-Gal, the released galactose is not trapped in the active site (Boix *et al.*, 2002); this is consistent with the more open structure of active site in the mutant enzyme resulting from the Arg365Lys mutation.

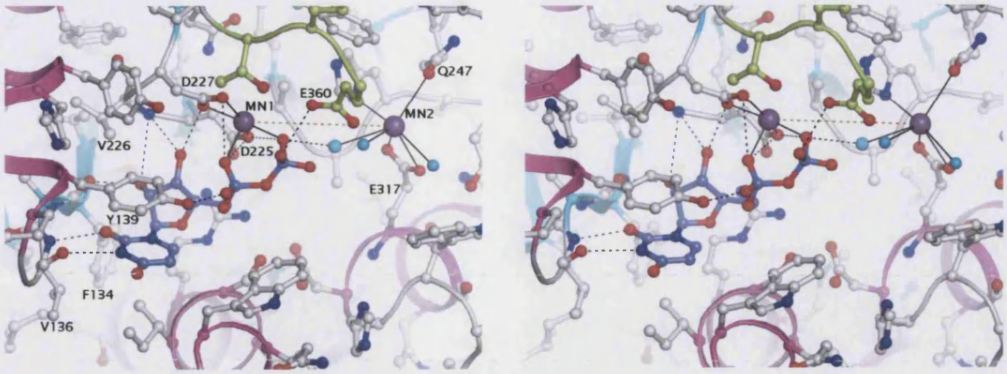


Figure 3.7: Stereo diagram showing the active site environment of the Arg365Lys mutant in complex with UDP and 2 manganese ions shown as purple and violet spheres. The catalytic manganese is designated as MN1 while the second manganese is designated as MN2. UDP is shown in slate and the C-terminal residues are shown in light green. The water molecules making direct interactions with the second manganese are shown as cyan spheres. The red dashed line represents the distance between the catalytic manganese (MN1) and the second manganese (MN2) which is approximately 7.5Å.

Structure of the Arg365Lys complex with UDP-2F-Gal

Crystals of the Arg365Lys mutant were grown in the presence of UDP-2F-Gal, manganese and GlcNAc (Table 3.1). The presence of GlcNAc was found to be essential for crystallisation, but no electron density corresponding to a GlcNAc molecule was identified in the structure. The space group was triclinic with 2 molecules per asymmetric unit and the structure was solved at a resolution of 2.4Å. In the structure, an intact UDP-2F-Gal molecule and one manganese ion are bound in the active site. Two regions of structure show varying levels of disorder in other α3GT structures, residues 190-195, centred around Trp195 and residues 358-368 at the C-terminus (Figure 3.3). In this complex, the Trp195 loop is ordered and in a different conformation from that in apo-forms of wild-type and Arg365Lys α3GT while, also unlike the apo-enzyme structures, the C-terminus is partly ordered (Figure 3.3B). In molecule B, residues Thr358, Lys359, Glu360, Tyr361 and Asn362 are ordered, Asn363 is modelled as a glycine because of the absence of any defined density for the side chain and residues 363-368 are disordered. In molecule A, residues Thr358 and Lys359 are ordered and residues 360-368 are disordered,

suggesting that the greater order in molecule B may result from molecular contacts in the crystal. The partially ordered C-terminal structure appears to be stabilized by interactions between Lys359 and the β-phosphate of the inhibitor. It seems likely that two factors may prevent this region from forming the highly structured closed conformation observed in the UDP complex of wild-type α3GT. One is the presence of the 2F-Gal (2F-Gal) attached to the β-phosphate of the UDP that appears to displace the network of water molecules which stabilize the interactions between residue 365 and the β-phosphate that appear to be important in stabilizing the closed structure in the UDP complex (Boix *et al.*, 2001). Secondly, the mutation of Arg365 to Lys may compromise interactions that are required for the stabilization of the closed conformation. Figures 3.8 and 3.9 show the structure of the active site around the bound UDP-2F-Gal. The UDP moiety is bound similarly to UDP in previously characterized α3GT complexes, except that there are fewer interactions between residues in the C-terminal region and the diphosphate; most notably, there are no interactions involving Tyr361 and Lys365, the site of the mutation. The inhibitor is bound in a distorted configuration in which the 2F-galactosyl moiety is bent back almost parallel to the plane of the diphosphate, a conformation that is similar to that of UDP-2F-Gal bound to the bacterial retaining galactosyltranferase, LgtC (Persson *et al.*, 2001). As shown in Figure 3.12, the structure is different from the structure of the Hg-UDP-Galactose in the form I structure of Gastinel and co-workers (Gastinel *et al.*, 2001). The pyranose ring adopts a more distorted structure than the more stable $^4C_1$ chair conformation often observed in other UDP-Gal containing complexes in the RCSB-Protein Data Bank; the O5 of the ring puckers forward in such a way that it disrupts the stable chair conformation. Key residues responsible for stabilizing the 2F-Gal in this conformation include several residues that are highly conserved in homologues of α3GT: Arg202, which is adjacent to the Trp195 loop, Asp225 of the Asp225-Val-Asp227 motif associated with manganese binding, and His315, Asp316 and Glu317. Also His280, which varies among α3GT homologues and corresponds to Leu/Met266 that has a major role in donor substrate specificity in the histo-blood group A and B transferases, interacts with the fluorine atom at 2′ position (Table 3.2).

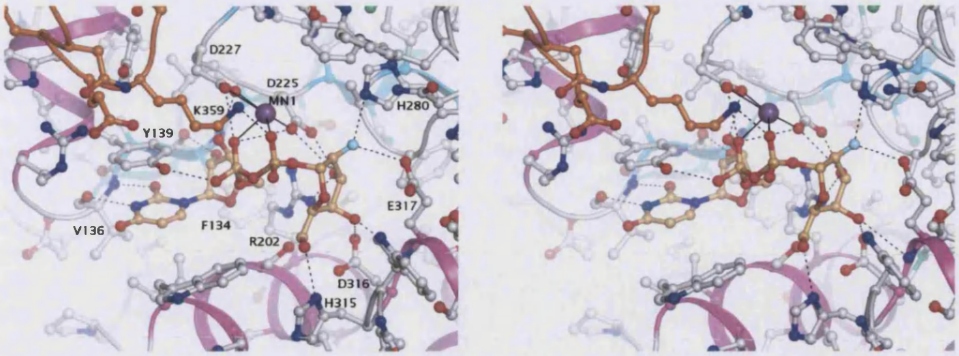


Figure 3.8: Stereo diagram showing the active site environment of the Arg365Lys mutant in complex with the catalytic manganese and donor analogue, UDP-2F-Gal (bright orange). Residues from the C-terminal region are shown in orange.

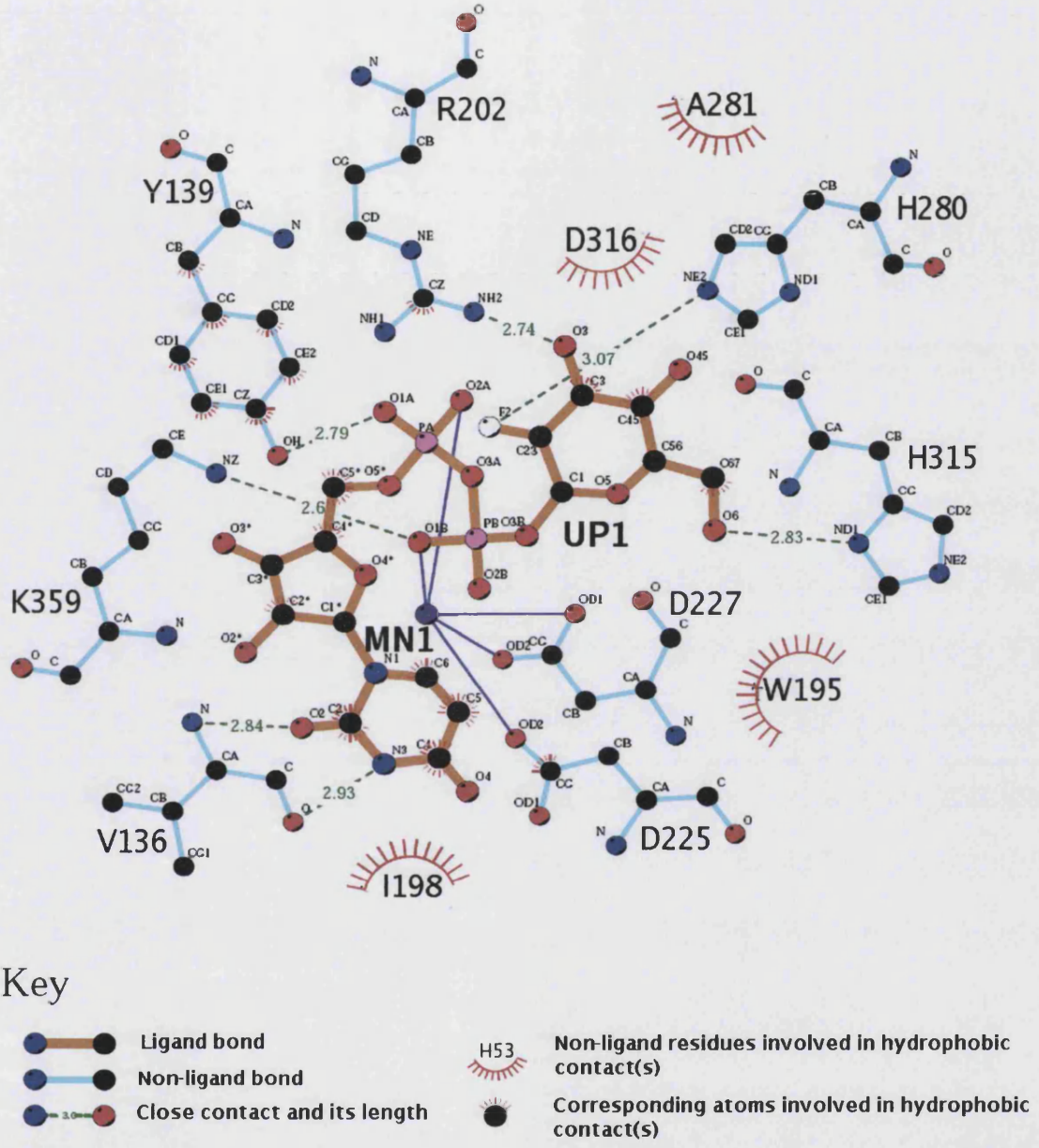


Figure 3.9: Schematic diagram showing the main contact residues including hydrogen bond and hydrophobic interactions calculated by the program HBPLUS(McDonald and Thornton, 1994) between the residues of Arg365Lys mutant and the donor analogue, UDP-2F-Gal.

Table 3.2: Close atom contacts (< 3.5 Å) between the 2F-Gal moiety of UDP-2F-Gal and α3GT mutant Arg365Lys

Ligand atom (UDP-2F-galactose)	*Interacting atoms (α3GT- Arg365Lys)*	*Distance (Å)*
F2'	NE2 280 His	3.00
	OE2 317 Glu	3.20
O3'	OD1 225 Asp	2.74
	NH2 202 Arg	2.74
	OD2 316 Asp	3.30
O4'	N 317 Glu	3.20
	OD2 316 Asp	2.79
O6'	ND1 315 His	2.83

van der Waals contacts

Residue	*Number of contacts*	*Residue*	*Number of contacts*
Phe134	3	Val226	1
Ala135	4	Asp227	4
Val136	4	His280	2
Tyr139	32	Ala281	8
Trp195	4	His315	3
Ile198	6	Asp316	6
Ser199	4	Glu317	3
Arg202	7	Lys359	2
Asp225	8		

Contact distances are the following maximum allowed values: C-C, 4.1 Å; C-N, 3.8 Å; C-O, 3.7 Å; O-O, 3.3 Å; O-N, 3.4 Å; N-N, 3.4 Å; C-S, 4.1 Å; O-S, 3.7 Å; N-S, 3.8 Å; C-F, 3.17 Å; N-F, 3.03 Å. The contact distances were calculated using CONTACT from the CCP4 suite(CCP4, 1994).

C-terminal truncations and Lys359 mutations

To further explore the roles of residues in the flexible C-terminus region, and conformational changes involving this region, we have investigated the effects of progressive C-terminal truncations and substitutions for Lys359 on catalytic activity. Lys359 was selected because, like Arg365, it is totally conserved in all known relatives of α3GT. Studies carried out by Tumbale and co workers (Jamaluddin *et al.*, 2007) showed that the truncation of up to three amino acids from the C-terminus (CΔ3 mutant) had little effect on activity but removal of a fourth residue (Arg365) produced a major loss in activity arising from a 120-fold reduction in k_{cat} and smaller changes in the K_m and K_i for UDP-Gal. It should be noted that this result is in conflict with a previous report by Henion *et al.* who found that the removal of just three residues of the C-terminus resulted in a complete loss of activity (Henion *et al.*, 1994).

Crystals of the CΔ3 mutant were grown in the presence of UDP-Gal and *N*-acetyl lactosamine (LacNAc); their space group is C2 with one molecule per asymmetric unit. The structure, which was solved at 2.0 Å resolution, contains a manganese ion and UDP in the active site. An additional region of electron density was, on refinement, identified as a Tris ion bound in the active site. This ion appears to disrupt interactions of Lys359 and Arg365 with the diphosphate moiety of the UDP but eight C-terminal residues (residues 358-365) are highly ordered, apparently being stabilized in part by interactions between Glu360 and the Tris. The Tris ion is bound at a location similar to that of the second manganese ion in the Arg365Lys•UDP complex and has similar structural effects in disrupting some interactions of the C-terminus with the UDP and the main body of the enzyme. Both ions are located at a site anticipated to overlap that of the trisaccharide product. The conformation of the C-terminal region in CΔ3 is distinct from that in the form II structure of wild-type α3GT (Figure 3.3B and 3.10). It is possible that the absence of the last three residues, in particular Asn367 and Val368, which interact with the Trp195 loop in the wild-type enzyme, may contribute to the change in structure in this mutant. However, the components of the binding site for the acceptor substrate (Boix *et al.*, 2002) are intact

and the functional properties of the truncated enzyme show only minor changes (Jamaluddin *et al.*, 2007).

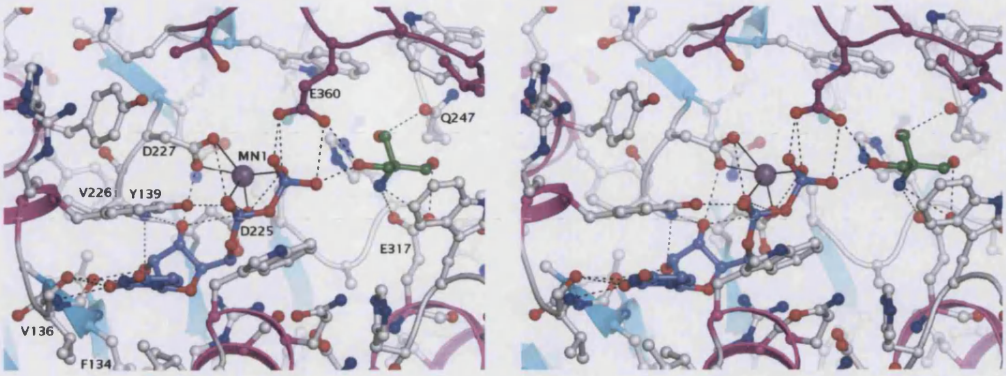


Figure 3.10: Stereo diagram showing the active site environment of the CΔ3 mutant in complex with UDP (slate) and Tris ion (forest green). C-terminal residues are represented in purple while the catalytic manganese is shown as a purpleviolet sphere. Interatomic distances between 2.5Å – 3.5Å are shown as black dashed lines while the black solid lines represent the metal coordinating interactions.

The flexible C-terminal region includes two basic amino acids that are totally conserved in all known members of the α3GT family, Lys359 and Arg365. In previous structures, both Lys359 and Arg365 have been observed to interact directly with phosphates of the bound UDP and acceptor substrates (Boix *et al.*, 2002; Boix *et al.*, 2001) and the role of Arg365 has previously been investigated through the properties of the Lys mutant (Zhang *et al.*, 2003). α3GT variants with conservative (Arg) and side chain truncated (Ala) substitutions for Lys359 were therefore expressed to investigate the role of this residue. Studies by Tumbale and co workers also showed that the principal effects of both of these changes were to reduce k_{cat} : 34-fold for the Arg substitution and 350-fold for the Ala substitution (Jamaluddin *et al.*, 2007). Substitutions for both Arg365 and Lys359, like the deletion of Arg365, therefore lower transition state stability. This is consistent with their conservation in the related enzymes which differ in specificity for donor and acceptor substrates, suggesting that they are important for catalysis and/or structure rather than substrate specificity. It is interesting to note that attempts at co-crystallising the Lys359Arg mutant with UDP-Gal resulted in a structure of the mutant in complex with just UDP,

manganese and glycerol bound in the active site. Although the mutant was crystallised in the presence of UDP-Gal, only the UDP moiety is seen to be bound in the active site. The cell dimensions of the crystal and overall structure of the mutant is similar to that of previous structures of Arg365Lys mutant in complex with UDP and 2 manganese ions and the structure of the CΔ3 in complex with UDP and Tris ion. The Lys359Arg mutant showed a glycerol molecule bound in a similar position to the second manganese ion and the tris molecule found in the Arg365Lys•UDP and CΔ3•UDP complexes (Figure 3.11). The mutation of Lys359 to Arg could not be observed as the C-terminal region is disordered. Refinement statistics for the structure of Lys359Arg mutant in complex with UDP and glycerol are listed in table 3.3.

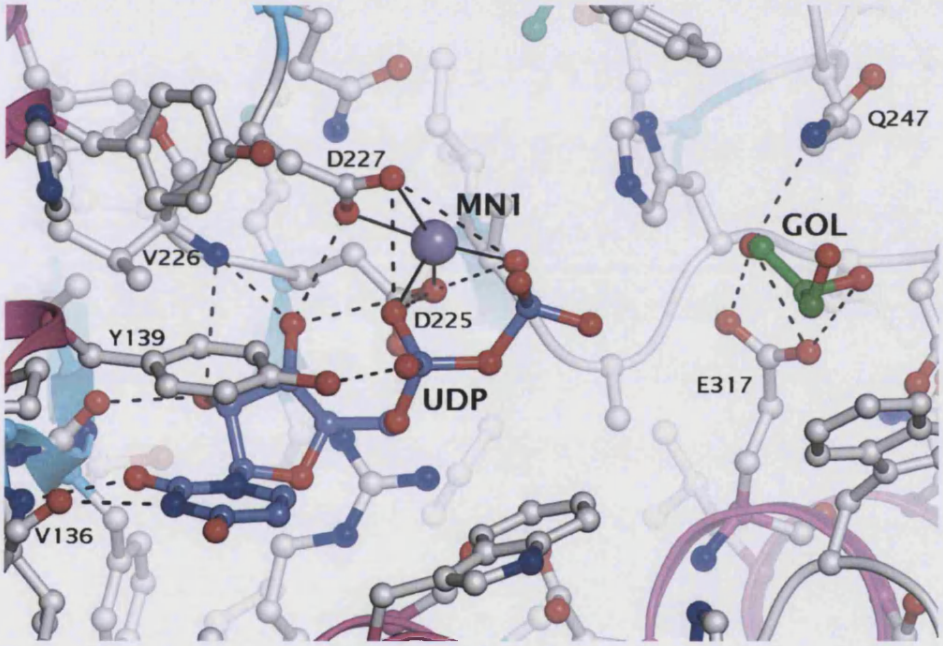


Figure 3.11: Stereo diagram showing the active site environment of the Lys359Arg mutant in complex with UDP (slate), glycerol (green) and catalytic manganese is shown as a purpleviolet sphere (MN1). Interatomic distances between 2.5Å – 3.5Å are shown as black dashed lines while the black solid lines represent the metal coordinating interactions.

Table 3.3: Data collection and refinement statistics for Lys359Arg-UDP complex structure

Substrates observed in the crystal structure	Manganese UDP moiety Glycerol molecule
Disordered regions	C-terminus (358-368)
Space group	C2
Mols/A.S.U	1
Crystal parameters	
a (Å)	123.19
b (Å)	68.61
c (Å)	44.83
α (°)	90.0
β (°)	89.66
γ (°)	90.0
Resolution (Å)	50.0-2.10
R_{cryst}[c]/R_{free}[d] (%)	18.7/22.6
Ramachandran Plot	
% core/allowed	89.9 /10.1
***B*-factor statistics (Å^2)**	
Protein[e]	12.08 (A)
Ligand in the active site[e]	10.3 (UDP), 41.4 (GOL)
Metal	11.7 (MN1)

a $R_{symm}=\Sigma_h\Sigma_i[|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)]$, where *Ii* is the *i*th measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$.
b Outermost shell are 2.15-2.10 Å .
c $R_{cryst}= \Sigma_h|F_o - F_c|/\Sigma hF_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection *h*.
d R_{free} is equal to R_{cryst} for a randomly selected 5.1% subset of reflections, not used in refinement.
e Temperature factors for individual molecules quoted

Discussion

The structure of the UDP-2F-Gal complex shows a conformation for the donor substrate analogue that contrasts with the configuration of UDP-Gal in the earlier structure of Gastinel *et al.* (2001) of a complex containing Hg-UDP-Gal. The UDP moiety interacts with α3GT similarly to the UDP in this and other previously-described structures, while the 2F-Gal is bound in a bent, apparently strained configuration similar to that of the Hg-UDP-Gal structure (Gastinel *et al.*, 2001). An

interesting feature of this structure is the deformation of the pyranose ring of 2F-Gal which could be a result of the substitution of oxygen for fluorine at the 2′ position. Fluorine is much more electronegative than a hydroxyl group and therefore would produce much stronger electrostatic interactions with both NE2 of His280 and OE2 of Glu317 which results in a change in the overall puckering of the ring. The hydroxyl groups circling the 2F-Gal pyranose ring coordinate interactions with the surrounding amino acid residues hence giving insight into key residues that may play a crucial role in catalysis and substrate specificity. The residues that make close contacts (< 3.5Å) with the 2F-Gal ring include Glu317, Asp316, Asp225, His280, His315 and Arg202 with the last two residues making direct hydrogen bond interactions with the sugar ring. Interestingly the orientation of the UDP moiety of Hg-UDP-Gal in the wild type structure (Gastinel *et al.*, 2001) is different from that of UDP-2F-Gal in the Arg365Lys mutant structure, with the biggest difference observed in the position of the uracil ring of UDP (Figure 3.12). The orientation of UDP from the Hg-UDP-Gal is influenced by the binding of the mercury (Hg^{2+}) ion close to the uracil ring thereby shifting its position to minimise steric hindrance.

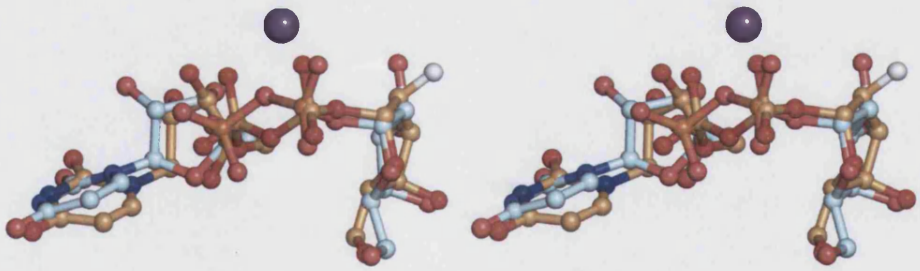


Figure 3.12: Stereo view representation of the Hg-UDP-Gal from the form I structure (Gastinel *et al*, 2001; PDB code: 1G93) shown in light cyan, superimposed on top of the UDP-2F-Gal from the Arg365Lys mutant structure shown in bright orange. Manganese ion is shown as a purple violet sphere.

Furthermore, the presence of the mercury ion might contribute to the disorder of the C-terminal region observed in the wild type structure. The Arg365Lys•UDP-

2F-gal complex is a plausible structure for a donor substrate-enzyme complex that suggests a role for ground-state destabilization of the UDP-Gal in the catalysis of galactose transfer to disaccharide acceptors and other molecules including water (Zhang *et al.*, 2001) and azide (Monegal and Planas, 2006). Interestingly the low catalytic activity of α3GT for galactose transfer to water or azide is an inverting reaction, producing β-Galactose and β-azido galactose in contrast to the retaining reaction for transfer to larger oligosaccharide substrates (Boix *et al.*, 2002; Monegal and Planas, 2006). We have used the structures of the UDP-2F-Gal complex and UDP plus lactose complex to model a ternary complex containing both UDP-2F-Gal and lactose. This suggests that there is no steric clash when both substrates bind to the enzyme which is consistent with kinetic studies that show α3GT utilizes a sequential mechanism in which UDP is not released prior to acceptor substrate binding (Boix *et al.*, 2002). The interactions between the 2F-Gal and enzyme are summarized in Table 3.2; they are different from those between the wild-type enzyme and β-Galactose in the structure of the complex containing hydrolyzed UDP-Gal complex (Boix *et al.*, 2002) (Figure 3.13). The Arg365Lys mutant retains significant catalytic activity, it is reasonable to suggest that the complex of UDP-2F-Gal complex with this mutant provides a suitable model for the α3GT•UDP-Gal complex. Consequently, residues making H-bond interactions with the 2F-Gal are likely to play important roles in specificity for the UDP-Gal relative to other UDP-sugars. It is interesting that in the histo-blood group A and B synthases, a *N*-acetylgalactosaminyl-transferase and a galactosyltransferase, respectively, that differ in sequence at only 4 sites, a key residue that contributes to their difference in specificity is Leu266 or Met266, of the A and B transferases, respectively, which corresponds to His280 of α3GT (Patenaude *et al.*, 2002). The presence of this residue may be a key to the high specificity of α3GT for UDP-Gal *vs* UDP-GalNAc since the imidazole ring appears to block the accommodation of the more bulky acetamido group on C2 of the galactose ring. The fact that mutation of His280 to Ala produces an enzyme with a low but significant level of UDP-GalNAc transferase activity is consistent with this view (Zhang *et al.*, 2003). Interactions of UDP-2F-Gal with Trp195 and Lys359 can account for the stabilization of the Trp195 loop and the C-terminus in the complex as compared with

the apo-enzyme. We suggest that in the complex of wild-type α3GT with UDP-2F-Gal, the presence of Arg, as opposed to Lys at position 365 may result in stronger interactions between ligand and the C-terminal region producing increased order in this region than in the Arg365Lys mutant. A conformational difference between the complexes of wild-type α3GT and the Arg365Lys mutant with the inhibitor may help to explain our ability to obtain crystals of the complex of UDP-2F-Gal with only the Arg365Lys mutant. In the UDP complex of wild-type α3GT, the C-terminus is highly ordered as compared with the UDP-2F-Gal complex of the Arg365Lys mutant. There Arg365Lys mutant complex with UDP is also somewhat more ordered than the UDP-2F-Gal complex; however, this structure is complicated by the presence of a second manganese ion.

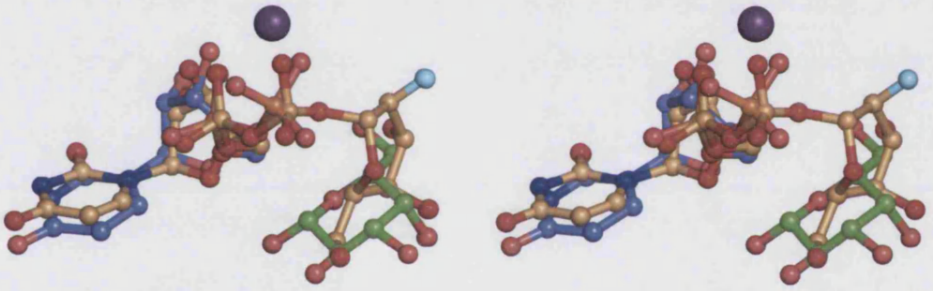


Figure 3.13: Stereo view of the UDP and galactose from the wild type α3GT structure (Boix *et al,* 2001; PDB code: 1GX0) shown in slate and green colour respectively, superimposed on top of UDP-2F-Gal from the Arg365Lys mutant structure shown in bright orange. Manganese ion is shown as a purple violet sphere.

In the histo-blood group A and B glycosyltransferases, there are flexible loops corresponding to those in α3GT. Unlike α3GT, these loops are disordered in enzyme complexes containing UDP and acceptor substrates as well as in the apo-enzymes (Patenaude *et al.*, 2002). However naturally-occurring mutations in both loops have been shown to affect enzyme activity. Interestingly, the mutations in the large (Trp195) loop specifically affect B-transferase activity while mutations in the C-

terminal region affecting the Lys and Arg corresponding to Lys359 and Arg365 of α3GT affect the activities of both A and B transferases (Yazer and Palcic, 2005). Therefore, although the two loops in the A and B transferases do not become stabilized in the presence of UDP, with or without acceptor substrate, in contrast to α3GT, the effects of mutations suggest that these flexible regions play a significant role in the enzyme mechanism. This may reflect interactions between Trp195 and the Gal or GalNAc of the donor substrate that do not occur in the UDP complex. Structural changes may occur in these regions only in the presence of the intact donor substrate, or in a ternary complex.

No structures are available for complexes containing both a UDP-Gal analogue and acceptor substrate or the UDP and trisaccharide product. However, as discussed previously, we have determined the structure of complexes containing a cleaved UDP-Gal, i.e. the products of galactose transfer to water, and of UDP plus acceptor substrates. The latter are dead-end inhibitory complexes while the former can be viewed as an enzyme•product complex of the UDP-Gal hydrolase activity. In all of these complexes, the C-terminus is highly structured, as in the UDP complex, being stabilized by interactions between Tyr361 and the β-phosphate and by direct and solvent mediated interactions between Arg365 (Lys in the Arg365Lys mutant) and both phosphates. A significant distinct feature of the closed conformation is a network of solvent molecules that is conserved between the various complexes that have this type of structure; these solvent molecules mediate interactions that stabilize the C-terminus. The presence of the 2F-Gal blocks the formation of this network, maintaining a more open conformation. Catalysis appears to be linked to a transition between this open conformation in which the C-terminus is partially ordered and the closed conformation present in the previously described structure of wild-type enzyme containing UDP plus galactose. The selective reduction in k_{cat} arising from Lys359 and Arg365 mutations and the deletion of residues 365-368 are consistent with this being important for catalysis.

There is growing interest in the role of protein dynamics in protein functions including catalysis (Daniel *et al.*, 2003; Hammes-Schiffer and Benkovic, 2006). GTs

can be divided into 65 or more families based on functional properties and sequence inter-relationships, but a limited array of them have been structurally characterized. The presence of one or two loops that are disordered in the absence of ligands but develop distinct ordered structures on substrate binding appears to be a widespread feature of these enzymes (Unligil and Rini, 2000). α3GT complexes containing UDP or UDP-2F-Gal show increased order in these loops relative to apo-enzyme structures but also much lower global temperature factors (17 $Å^2$ vs 60 $Å^2$). This may arise, in part, due to differences in molecular contacts in the different crystal forms but also appears to reflect the conformational flexibility of the ligand-free form.

The presence of a second manganese in the active site in the UDP complex of the Arg365Lys mutant, as well as the more open conformation, distinguishes its structure from that of the previously characterized complex with UDP (Boix *et al.*, 2001). The second manganese, like the Tris in the similar CΔ3 mutant structure, interacts with residues involved in binding acceptor substrates (Glu317, Gln247) and the galactose moiety of UDP-Gal (His280). The binding of a second cation in the active site has not been previously observed in complexes of the wild-type enzyme and it is possible that they could be artefacts linked to the mutations which perturb interactions between the C-terminus, ligands and active site. However, previous kinetic studies have shown that the activity of α3GT is modulated by the binding of manganese ions to two sites, a high affinity site with a K_d of 80 μM whose occupancy is essential for activity and a second, weaker site (K_d of 2 mM), that increases V_m about 3-fold (Zhang *et al.*, 2001). It is possible that the weaker binding site is that occupied by the second manganese in the Arg365Lys complex with UDP (and is also mimicked by the Tris in the CΔ3 complex with UDP). The structure indicates that the second manganese can only bind after galactose has been transferred to the acceptor substrate and the trisaccharide product has dissociated from the enzyme. A plausible explanation of the effects of the two manganese ions is that the tight binding manganese is directly involved in UDP-Gal binding and catalysis, the second ion increases k_{cat} by facilitating the disruption of interactions between the C-terminal loop and UDP product, facilitating the dissociation of UDP, the final step in the catalytic cycle.

Conclusion

The results presented here show that the binding of a donor substrate analogue induces conformational changes in both the ligand and enzyme. Two loops of α3GT are stabilized in the complex of which the C-terminal region, in particular, is highly flexible. Structural transitions in this region are connected with donor substrate binding and distortion (ground state destabilization), cleavage of the UDP to galactose bond, formation of a binding site for acceptor substrate and UDP release. Both mutational and structural studies accentuate the key roles of two highly conserved basic residues, Lys359 and Arg365 in both C-terminal structural changes and catalysis.

Publication:

Five of the structures presented in this chapter have been published in the Journal of Molecular Biology. The coordinates of all five structures have been deposited to the Protein Data Bank (PDB) and their codes can be obtained from the reference below:

Jamaluddin, H., Tumbale, P., Withers, S.G., Acharya, K.R. and Brew, K. (2007) Conformational Changes Induced by Binding UDP-2F-Galactose to alpha-1,3 Galactosyltransferase- Implications for Catalysis. *J. Mol. Biol.* **369**: 1270-1281.

3.3 In search for the catalytic nucleophile of α-1,3 Galactosyltransferase (α3GT)

3.3.1 Introduction

One of the main aims of this project was to shed light into the structural basis of α3GT catalytic mechanism. In particular to try to search for the ever elusive catalytic nucleophile which was implicated in a double displacement mechanism first, proposed by Gastinel and co-workers in 2001. Based on the form I structure (Gastinel *et al.*, 2001), Glu317 was purported to be the catalytic nucleophile, however, the low resolution of the structure and high *B*-factors have cast doubts on this conclusion. In 2003, Zhang and co-workers reported that mutation of Glu317 to Gln affected the rate of galactosyltransfer however, the magnitude of this change does not support the proposal that Glu317 is the catalytic nucleophile in a double displacement mechanism (Zhang *et al.*, 2003). We have carried out crystallisation attempts of the Glu317Gln mutant in complex with substrates in order to obtain conclusive evidence in support of the double displacement mechanism. Therefore, this chapter describes the crystallisation, structure determination, refinement and analysis of the first structure of a mutant of α3GT with an intact active donor substrate bound in the active site. This chapter will also discuss how this Glu317Gln-UDP-Gal complex structure has led to further mutational and structural studies and the implications that these mutant structures have towards the understanding of catalysis and specificity of α3GT.

3.3.2 Glu317Gln mutant in complex with UDP-Gal

Crystallisation

We sought to obtain the crystal structure of the Glu317Gln mutant in complex with UDP-Gal. The protein was prepared as previously described in subchapter 3.1. 10mM $MnCl_2$ and 10mM UDP-Gal were then added to the protein sample and incubated on ice for about 1 hour prior to crystallisation. Initial crystallisation attempts were carried out based on the published crystallisation condition which consisted of 5-10% PEG 6000 and 0.1M Tris-HCl pH 8 (Boix *et al.*, 2002; Boix *et*

al., 2001; Zhang *et al.*, 2004; Zhang *et al.*, 2003). However, this condition did not produce any co-crystals but instead caused thick precipitation. An additive screen based on the original crystallisation condition was then carried out as described in subchapter 3.1. Among all the additives tested, the addition of MPD was observed to facilitate solubility of the protein and promote the formation of crystals (Table 3.0). Crystals of the Glu317Gln mutant in complex with UDP-Gal were grown at 16°C by vapour diffusion, by the hanging drop method by mixing 1 µl of the protein at 5mg/ml in 20mM MES-NaOH buffer, pH 6.0, 10% glycerol, containing 10 mM $MnCl_2$ and 10 mM UDP-Gal, with an equal volume of reservoir solution containing 10% PEG 6000, 0.1M Tris-HCl, pH 8.0 and 8% MPD. Single crystals grew to a relatively large size (Figure 3.14) and diffracted up to 1.82 Å (Figure 3.15). Crystallisation for the rest of α3GT mutants were carried out using this condition (PEG 6000 varying between 5-15% and MPD concentration varies between 2-20%) unless otherwise stated.

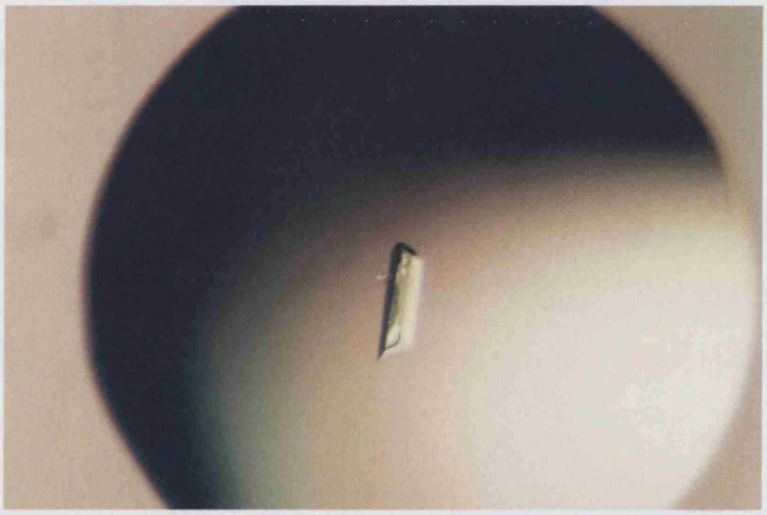


Figure 3.14: Crystals of Glu317Gln mutant in complex with UDP-Gal grew within 3-7 days at 19°C using the hanging drop technique at a protein concentration of 5 mg/ml. 1 µl of Glu317Gln mutant in 20mM MES-NaOH pH 6.0, 10% glycerol, 10mM $MnCl_2$ and 5mM UDP-Gal was mixed with 1 µl well solution containing 10% PEG 6000, 8% MPD and 0.1M Tris-HCl, pH 8. (crystal size: 0.5 X 0.2 X 0.1 mm)

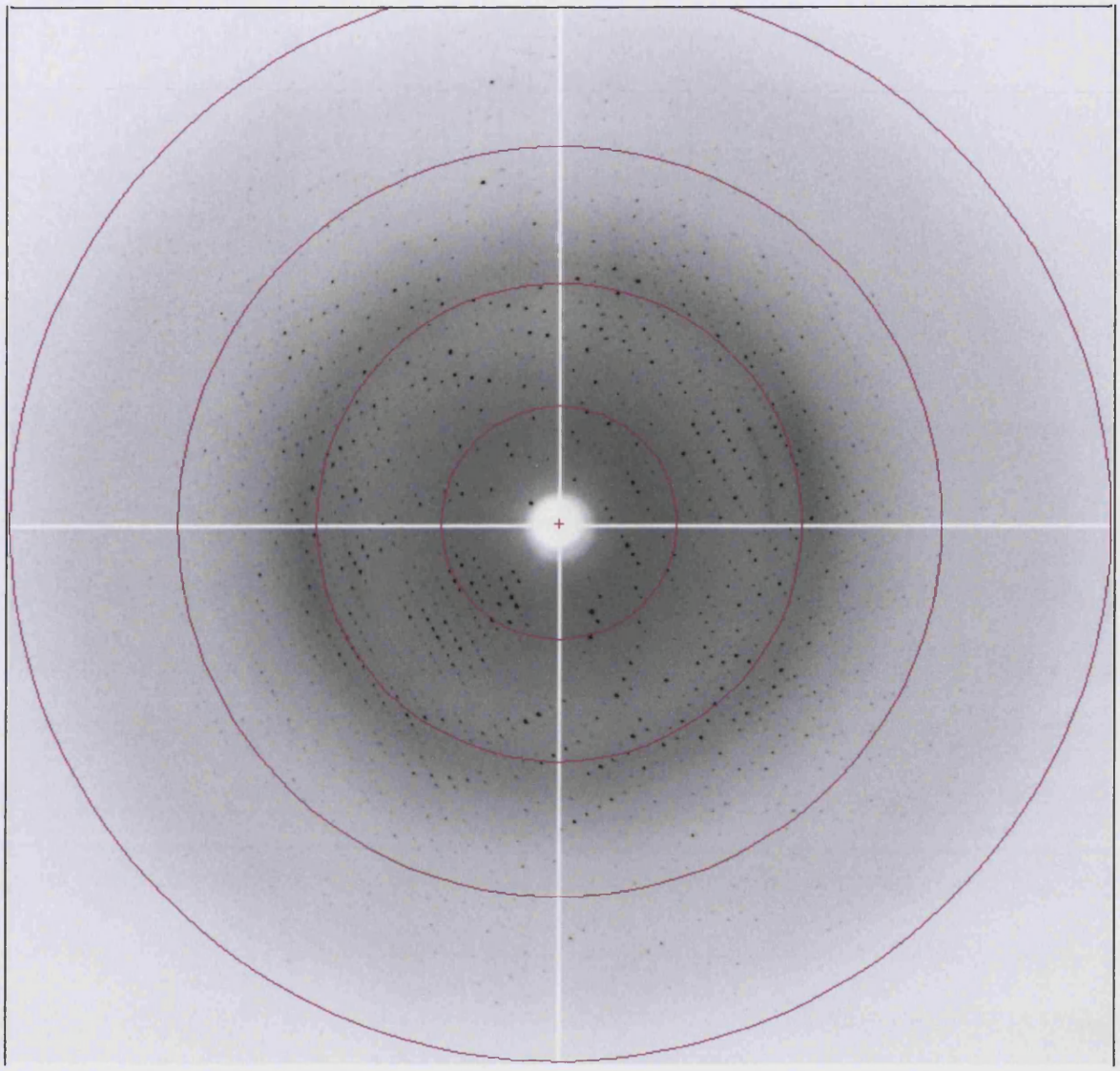


Figure 3.15: Diffraction image for the Glu317Gln mutant in complex with UDP-Gal. Collected on station 14.2 at the SRS, Daresbury, UK. The rings mark the resolution limits of 6 Å, 3 Å, 2 Å and 1.6 Å proceeding outwards from the centre. The data was processed to 1.82 Å

Data collection, Processing and Structure refinement

High resolution datasets were collected at the Synchrotron Radiation Source (SRS), Daresbury, UK for the Glu317Gln mutant in complex with UDP-Gal to 1.82 Å resolution at station PX 14.2, using an Area Detector Systems Corp. detector system. The data were collected at cryogenic temperature of 100K using a cryoprotectant of 25% glycerol. Raw data images were indexed and scaled using MOSFLM (Leslie, 2001) and SCALA (CCP4, 1994) respectively. The data were processed and scaled in the orthorhombic space group P222 to 1.82 Å. The cell dimensions were a=87.03 Å, b= 91.38 Å, c= 94.63 Å and $\alpha=\beta=\gamma=90^{o}$. The Glu317Gln co-crystal with UDP-Gal was highly mosaic but the high symmetry of the space group meant that relatively high completeness of 97.6% was still achieved (Table 3.4). The solvent content for the crystal was calculated using the CCP4 supported program, matthews_coeff. This helped to determine the number of molecules in the asymmetric unit. Calculations showed that for 2 molecules (molecular weight of ~34000 Da per molecule) of Glu317Gln mutant in the asymmetric unit, the Matthews coefficient was 2.77$Å^3$/Da with the solvent content of around 56%.

Although the crystal lattice was determined as primitive orthorhombic (P222), it still remained to be determined which of the subgroups the crystal belonged to. This was confirmed by looking at the systemic absences in each of the sub-space groups. Two criteria were chosen for a reflection to be deemed as present. Reflections that hold true to the general as well as special conditions specified for a particular sub-group should be present and the I/σI value for each reflection that conforms to the former criterion should be greater than 2. The dataset was processed in the space group P222 and molecular replacement was carried out for all possible sub-groups in P222 using the wild type form I structure, PDB code: 1K4V (Boix *et al.*, 2001) as a search model. Finding a correct solution by molecular replacement is usually indicative that the space group used is correct. Structure solution was carried out by molecular replacement using the program MOLREP (CCP4, 1994) to find 2 molecules in the asymmetric unit. The best solution was found in $P2_12_12_1$ space group with R-factor of 45.7% and correlation score of 63.8%. The dataset was also

checked against the space group determination program, POINTLESS (CCP4, 1994) and this program also picked up space group $P2_12_12_1$. Analysis of the crystal packing of the structure solution in $P2_12_12_1$ space group did not show any clashes between the dimer and also the symmetry related molecules which is indicative that the solution obtained was correct.

Crystallographic refinement was performed using the program package CNS (Brunger *et al.*, 1998) and model building was carried out using the program COOT (Emsley, 2004). Procedures carried out with CNS included the slowcool protocol of simulated annealing at 2500K using a maximum likelihood target function, restrained individual B-factor refinement, conjugate gradient minimisation, and bulk solvent correction. The behaviour of the R_{free} value [using a test set comprising of 1069 (1.6%) reflections] was monitored throughout refinement. Iterative cycles of refinement alternated with model building in COOT improved the quality of the structure and brought down the conventional (R_{cryst}) as well as the R_{free} values. Water molecules were gradually included into the model at positions corresponding to peaks in the $|F_o| - |F_c|$ electron density map with heights greater than 3σ and at H-bond distance from appropriate atoms. Residues with poor side-chain density were modelled as alanines and regions with very poor main chain electron density were excluded from refinement as not to bias the model. One bound MPD molecule from the crystallisation medium was identified and was included in the final stages of refinement. The final R_{cryst} and R_{free} for this structure are 18.7% and 20.1% respectively (Table 4). The program PROCHECK (Laskowski *et al.*, 1993) was used to assess the quality of the final structure. Analysis of the Ramachandran plot showed all the residues to be within allowed region of the plot (Figure 3.16).

Table 3.4: Data collection and refinement statistics for Glu317Gln-UDP-Gal complex structure

Substrates observed in the crystal structure	Manganese
	Intact UDP-Gal
Disordered regions	C-terminus (358-368) except for Arg365
Space group	$P2_12_12_1$
Mols/A.S.U	2
Crystal parameters	
a (Å)	87.03
b (Å)	91.38
c (Å)	94.63
α (°)	90.0
β (°)	90.0
γ (°)	90.0
Resolution (Å)	50.0-1.82
R_{symm} (%)[a] (outermost shell)[b]	10.1 (51.7)
Completeness (%) (outermost shell)[b]	97.9 (97.9)
***I/σI* (outermost shell)[b]**	7.8 (2.48)
Total reflections	808570
Unique reflections	68394
R_{cryst}[c]/R_{free}[d] (%)	18.7/20.1
Ramachandran Plot	
% core/allowed	89.5 /10.5
RMSD from ideal	
Bond angles (°)	1.481
Bond lengths (Å)	0.015
Number of water molecules	723
***B*-factor statistics (Å^2)**	
Protein[e]	18.5 (A), 18.0 (B)
Ligand in the active site[e]	22.0 (GDU-A), 17.0 (GDU-B)
Metal[e]	23.8 (MN1-A), 18.4 (MN1-B)

a $R_{symm}=\Sigma_h\Sigma_i[|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)]$, where *Ii* is the *i*th measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of *I*(*h*).
b Outermost shell are 1.89-1.82 Å .
c $R_{cryst}= \Sigma_h|F_o - F_c|/\Sigma hF_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection *h*.
d R_{free} is equal to R_{cryst} for a randomly selected 1.6% subset of reflections, not used in refinement.
e Temperature factors for individual molecules quoted

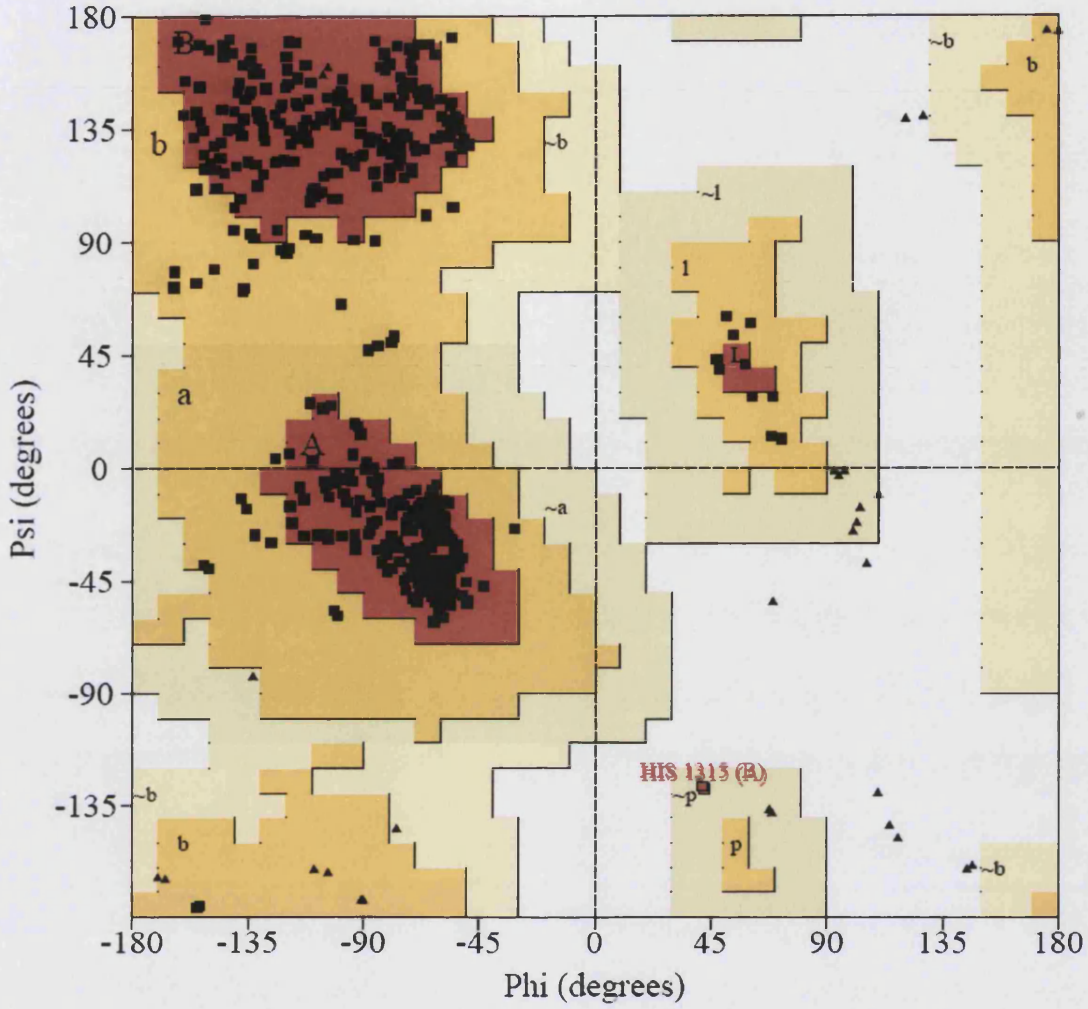


Figure 3.16: Ramachandran plot for the refined structure of Glu317Gln mutant of α3GT in complex with UDP-Gal at 1.82 Å resolution. The red coloured region of the plot (*A, B, L*) represents the residues in the most favoured regions; yellow regions (*a,b,l,p*) represents residues in the additional allowed regions; grey/buff coloured regions (*~a,~b,~l,~p*) represents residues in generously allowed regions and white areas of the plot represents the sterically disallowed regions. The glycine residues are represented by triangles; *A/a/~a* represent alpha helices; *B/b/~b* represent beta strands, *L,l,~l* represent left-handed helices; *p/~p* represent epsilon.

Results

The structure showed that the mutation produced no overall changes in the structure of the mutant as compared to the wild type structure. After the first round of refinement, the electron density showed unambiguous difference density corresponding to the manganese ion and full-length UDP-Gal in the active site. The subtle mutation from glutamate to glutamine could not be observed clearly from the electron density as the mutation is structurally conserved. The electron density for the overall protein was ordered and well defined however, electron density for residues that correspond to the c-terminal region (residues 358 to 368) was observed to be very disordered except for the side chain density of residue Arg365.

The previous structure of the wild type α3GT in complex with UDP-Gal showed that the galactose moiety was cleaved off and is situated a distance away (approximately 4 Å) from the β-phosphate of UDP, suggesting that hydrolysis of UDP-gal had taken place (Boix *et al.*, 2002). In this Glu317Gln mutant structure, a full length UDP-Gal is bound in the active site of each of the 2 molecules (Figure 3.17A) and this is consistent with kinetic studies which showed that this mutant has a markedly reduced hydrolase activity as compared to the wild type enzyme (Zhang *et al.*, 2003). The UDP-Gal is bound to the enzyme in a distorted conformation as opposed to a fully extended conformation observed for UDP-sugars in other proteins (Thoden and Holden, 1998). In this structure, the galactose moiety is bent almost perpendicular to the plane of the diphosphate region of UDP, suggesting a role for ground state destabilization in catalysis. The UDP moiety binds essentially in the same manner as in previous structures of the wild type and mutant α3GT bound with UDP (Boix *et al.*, 2002; Boix *et al.*, 2001; Zhang *et al.*, 2004; Zhang *et al.*, 2003). With the exception of C-terminal residues contacts, all other protein contacts with the UDP moiety are the same between the complexes. The galactosyl moiety of UDP-Gal is highly ordered within the active site of the Glu317Gln mutant as shown in the well-defined electron density for the intact UDP-Gal as shown in figure 3.17. The ring adopts the stable $^4C_1$ chair conformation and forms hydrogen bonds with the side chain atoms of residues Asp316, Asp225, His315, His280 and Arg202. Both O2' and O3' hydrogen bond with the side chain carboxylate of the Asp225 while

both O3' and O4' make hydrogen bonds with the carboxylate of Asp316 (Table 3.5), and this indicates that both these residues might play an important role in donor substrate binding and catalysis. Interestingly, Asp225 is invariant in the α1,3-glycosyltransferase family whilst Asp316 is conserved in all α3GalT orthologs (Gastinel *et al.*, 2001). Similar bidentate hydrogen bonding has also been observed in structures of other glycosyltransferases such as the retaining galactosyltransferase LgtC from *Neisseria meningitides* (Persson *et al.*, 2001), cyclodextrin glycosyltransferases and α-amylases (Brayer *et al.*, 2000). The main chain nitrogen atom of the mutated residue, 317 is at a close interacting distance (3.3 Å) with O4'of the galactose moiety. However, the side chain of the mutated residue does not make any direct interaction with the UDP-Gal as it is positioned as if it is stacking with the galactosyl ring of UDP-Gal (Figure 3.17A). A single water molecule is observed to interact directly with the galactosyl moiety of UDP-Gal in which it makes bidentate interaction with both O6' and O5' of galactose and a single interaction with O1B on the β-phosphate region of UDP (Figure 3.17A).

Table 3.5: Hydrogen bond interactions between UDP-Gal and α3GT mutant Glu317Gln

Ligand atom (UDP-Gal)	**Interacting atoms (α3GT- Glu317Gln)**	**Distance (Å)**
O2'	OD2 225 Asp	2.58
	NE2 280 His	3.05
O3'	OD1 225 Asp	2.97
	NH2 202 Arg	2.86
O4'	OD2 316 Asp	2.78
O6'	ND1 315 His	2.77
Ligand atom (UDP-Gal)	**Interacting atom**	**Distance (Å)**
O5'	WAT	3.01
O6'		2.59
O1B		2.65

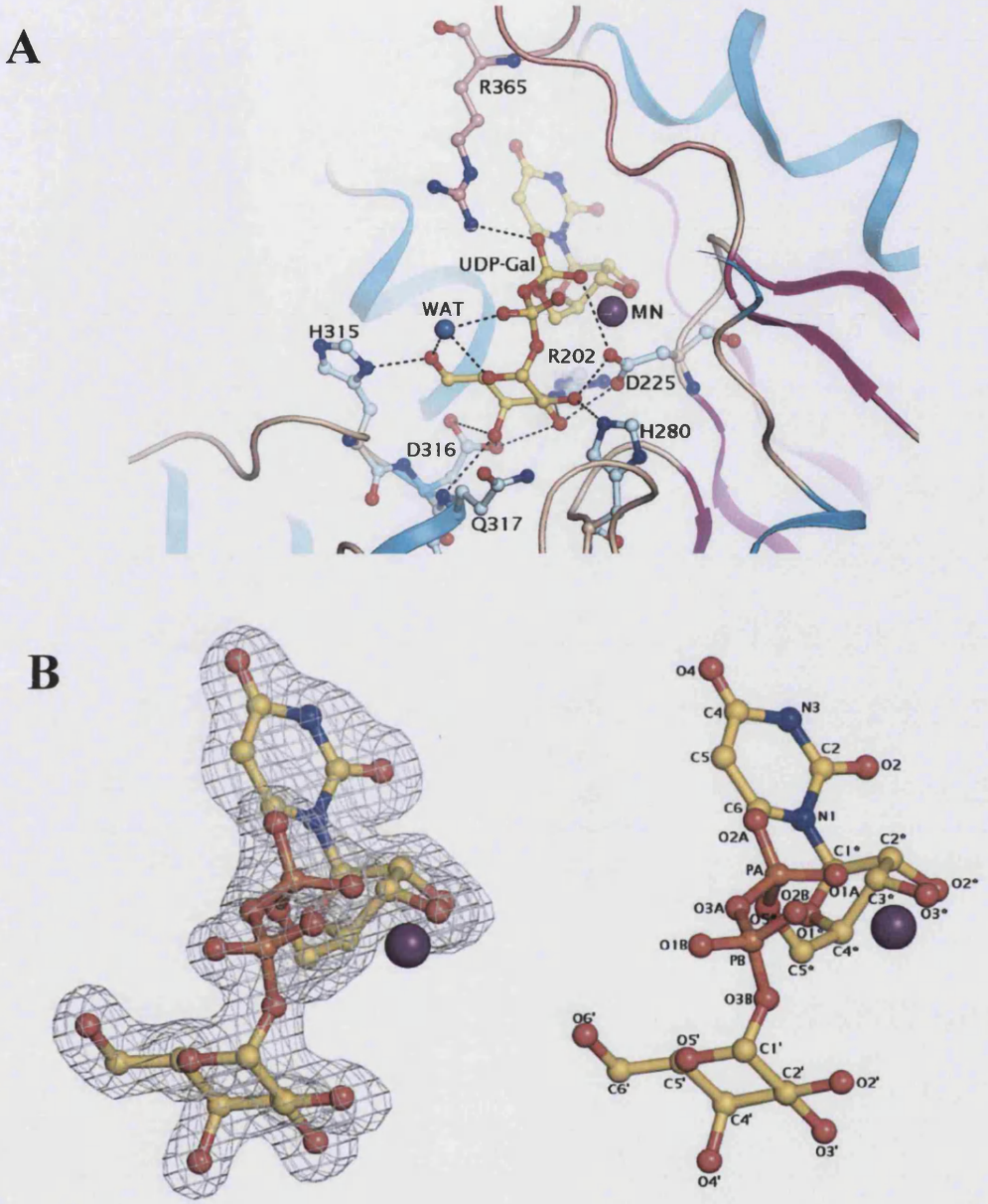


Figure 3.17: A- Close-up view of the active site interactions between the galactose moiety of UDP-Gal with the surrounding residues. Also shown are the manganese ion and a water molecule which interacts directly with the sugar. The UDP-Gal molecule is depicted in yellow, manganese ion (MN) as a purpleviolet sphere and water molecule as a blue sphere. Residues from the active site cleft are represented in lightcyan while the residue from the C-terminus is represented in salmon. Close interactions are represented as black dashed lines. B- The electron density in grey represents the F_o-F_c map, contoured at 3 σ, calculated from the final model, with UDP-Gal molecule omitted.

When this structure is compared to the structure of Arg365Lys mutant in complex with UDP-2F-Gal (Jamaluddin *et al.*, 2007), it is observed that this single water molecule hydrogen bonds with O5' of the galactose hence maintaining the stable $^4C_1$ chair conformation. It is reasonable to assume that the absence of a water molecule to interact with the galactosyl moiety of UDP-2F-gal results in a distorted chair conformation of the ring structure which along with the presence of a fluorine atom at position 2' of the ring, might contribute to the inhibitory nature of UDP-2F-Gal towards α3GT (Figure 3.18).

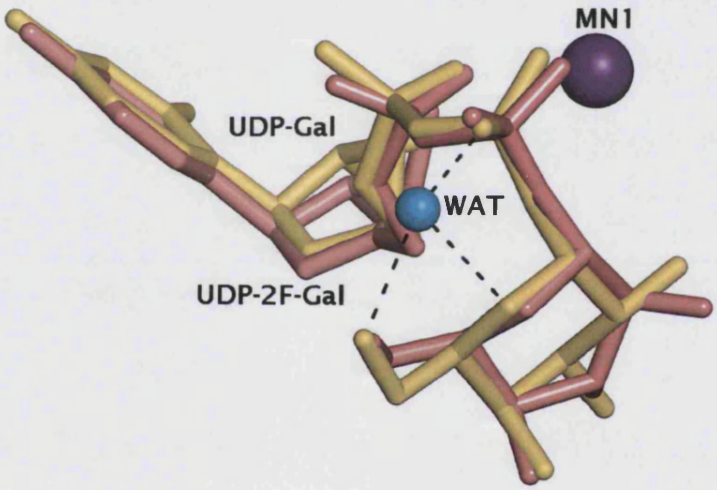


Figure 3.18: UDP-Gal from the Glu317Gln structure superposed with UDP-2F-Gal from the Arg365Lys structure (PDB code: 2JCF) (Jamaluddin *et al.*, 2007). The UDP-Gal is shown in yellow sticks while UDP-2F-Gal is shown in pink sticks. Manganese ion is shown in purple sphere while the water molecule which interacts directly UDP-Gal is shown in cyan. Close contacts are shown as black dashed lines.

Another interesting feature of this structure is that the C-terminus is mostly disordered except for the conserved residue Arg365. The side chain electron density for residue 365 is quite defined although the main chain density is virtually absent (Figure 3.19). The side chain of Arg365 interacts directly with the α-phosphate of the UDP moiety similar to the interactions observed in the wild type-UDP complex

structure. The difference in the interactions observed between Arg365 of this complex with the wild type-UDP complex is that it lacks the interaction with the network of water molecules surrounding the β-phosphate region of UDP. This lack of interaction is due to the presence of the galactose moiety which displaced the network of water molecules. It is predicted that interactions between Arg365 with these water molecules along with α-phosphate of UDP and Tyr139 function to stabilize the closed conformation of the C-terminus over the active site. The fact that only the side chain of Arg365 is observed to interact with the intact UDP-Gal while other residues on the C-terminus are disordered is indicative that Arg365 plays an important role in catalysis.

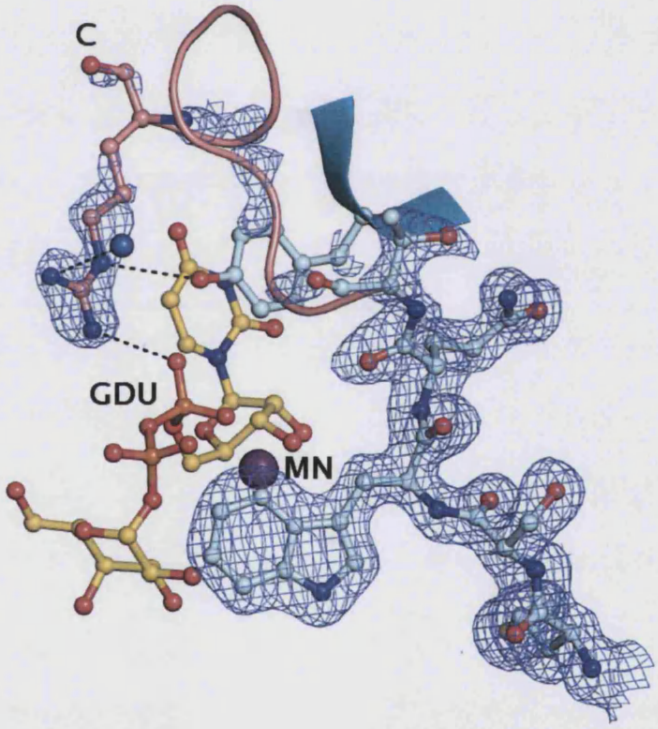


Figure 3.19: Shown is the C-terminal region (highlighted in salmon) of the Glu317Gln mutant in complex with UDP-Gal. Shown is the disordered and broken electron density for the C-terminus region compared to the rest of the enzyme (highlighted in lightcyan). The electron density in light blue represents the $2F_o$-F_c map, contoured at 1.0 σ, calculated from the first model after molecular replacement. Defined electron density is only observed for the side chain of residue Arg365.

We have also crystallised the Glu317Gln mutant in the presence of UDP-glucose in an attempt to obtain a complex structure of α3GT with an intact donor substrate (data not shown). However, in this structure, only the electron density for UDP moiety is visible while the glucose moiety is absent in the structure indicating that either the glucose moiety has been hydrolysed or that the glucose moiety is disordered due to the change in the vicinal hydroxyl chemistry at position 4'. If the former is true, our result would reflect another interesting finding on a mutant of bovine β4Gal-T1. In this study, mutation of residue 228, a residue in the vicinity of Glu317, from Arg to Lys increased the glucosyltransferase activity of β4Gal-T1 (Ramakrishnan *et al.*, 2005). It is thought that residue Glu317 in β4Gal-T1 (equivalent to Asp316 in bovine α3GT) is important in determining the low glucosyltransferase activity of the enzyme (Ramakrishnan *et al.*, 2005). Glu317 of α3GT is in close proximity to Asp316 hence mutation of this residue might explain the increased activity for UDP-glucose. However, detailed kinetic study on the Glu317Gln mutant of α3GT with UDP-glucose has yet to be carried out therefore, any conclusion on the effect of the mutation on glucosyltransferase activity could not be drawn out as yet.

Further structural studies on α3GT mutants based on the Glu317Gln complex structure

To further elucidate the catalytic mechanism of α3GT, several mutations were carried out in order to understand the role of the residues surrounding the active site in catalysis. The crystal structure of the Glu317Gln mutant allowed us to identify the residues which make direct interactions with the galactose moiety of UDP-Gal prior to hydrolysis and enabled further mutational studies to be carried out on these active site residues. Described here are the crystal structures of α3GT with Asp316 mutated to Asn and Glu in complex with UDP and LacNAc, His280Arg mutant, the AGGL mutant which has 4 mutations in residues which are predicted to be important in donor substrate specificity: His280Ala, Ala281Gly, Ala282Gly and Ile22Leu as well as the 5 site mutant which has mutations on 5 residues encompassing both the donor

substrate binding site as well as the acceptor substrate binding site; His280Gly, Cys338Leu, Gln247His, Ala248Pro and Trp249Gly.

3.3.3 Asp316 mutants in complex with UDP and N-acetyllactosamine (LacNAc)

Methods

In the crystal structure of the α3GT Glu317Gln, Asp316 is observed to interact directly with the O4' atom of the galactose moiety hence constructs of α3GT with Asp316 mutated to Asn and Glu were created in order to study the role of this residue in substrate binding and catalysis. Crystallisation of both these mutants was carried out in the presence of UDP-Gal and LacNAc using the same crystallisation methods described previously for the Glu317Gln mutant. The structures of both mutants were determined by the molecular replacement method using the native form II structure (PDB code: 1K4V (Boix *et al.*, 2001), as a search model with the program Molrep (Vagin and Teplyakov, 1997). Crystallographic refinement was performed using the program package CNS (Brunger *et al.*, 1998) for the Asp316Glu while crystallographic refinement was carried out using the program REFMAC for the Asp316Asn mutant structures. After the initial refinement, the difference electron density maps revealed the presence of all the mutated residues at their respective positions and UDP, manganese ion and LacNAc bound in the active sites of the Asp316Glu and Asp316Asn mutants. Several rounds of refinement and model building using the program COOT (Emsley, 2004) were performed until the R_{free} could not be improved. Water molecules were gradually included into the model at positions corresponding to peaks in the $|F_o| - |F_c|$ electron density map with heights greater than 3σ and at H-bond distance from appropriate atoms. Data collection and refinement statistics are shown in table 6. The program PROCHECK (Laskowski *et al.*, 1993) was used to assess the quality of the final structure. Analysis of the Ramachandran plot showed all the residues for the Asp316Glu and Asp316Asn complex structures to be within allowed region of the plot (Figure 3.20 and Figure 3.21 respectively).

"Table3. 6: Data collection and refinement statistics for Asp316Glu and Asp316Gln complexes

	Asp316Glu	**Asp316Asn**
Substrates observed in the crystal structure	Manganese UDP moiety LacNAc	Manganese UDP moiety LacNAc
Disordered regions	-	-
Space group	$P2_1$	$P2_1$
Mols/A.S.U	2	2
Crystal parameters		
a (Å)	45.25	45.12
b (Å)	94.63	94.38
c (Å)	94.69	94.94
α (°)	90.0	90.0
β (°)	98.9	99.3
γ (°)	90.0	90.0
Resolution (Å)	50.0-1.76	50.0-2.2
R_{symm} (%)[a] (outermost shell)[b]	7.9 (22.6)	12.0 (30.6)
Completeness (%) (outermost shell)[b]	93.1 (87.5)	94.9 (86.2)
***I/σI* (outermost shell)[b]**	20.36 (3.57)	9.74 (2.50)
Total reflections	412471	353562
Unique reflections	76955	39855
R_{cryst}[c]/R_{free}[d] (%)	18.3/21.0	18.3/22.7
Ramachandran Plot	87.8/12.2	90.5/9.5
% core/allowed		
RMSD from ideal		
Bond angles (°)	1.422	1.085
Bond lengths (Å)	0.007	0.009
Number of water molecules	1256	426
Ligand *B*-factor statistics (Å²)		
Protein[e] in the active site[e]	19.0 (A), 18.7 (B)	22.0 (A), 21.8 (B)
Metal[e]	15.6 (UDP-A), 15.0 (UDP-B) 17.1 (MN1-A), 16.7 (MN1-B) 25.0 (NAL-A), 25.6 (NAL-B)	17.6 (UDP-A), 16.9 (UDP-B) 19.2 (MN1-A), 19.4 (MN1-B)

a $R_{symm}=\Sigma_h\Sigma_i[I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)]$, where *Ii* is the *i*th measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of *I*(*h*).
b Outermost shell are 2.26-2.2 respectively .
c $R_{cryst}= \Sigma_h|F_o - F_c|/\Sigma hF_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection *h*.
d R_{free} is equal to R_{cryst} for a randomly selected are 1.3% and 3% subset of reflections respectively, not used in refinement.
e Temperature factors for individual molecules quoted.

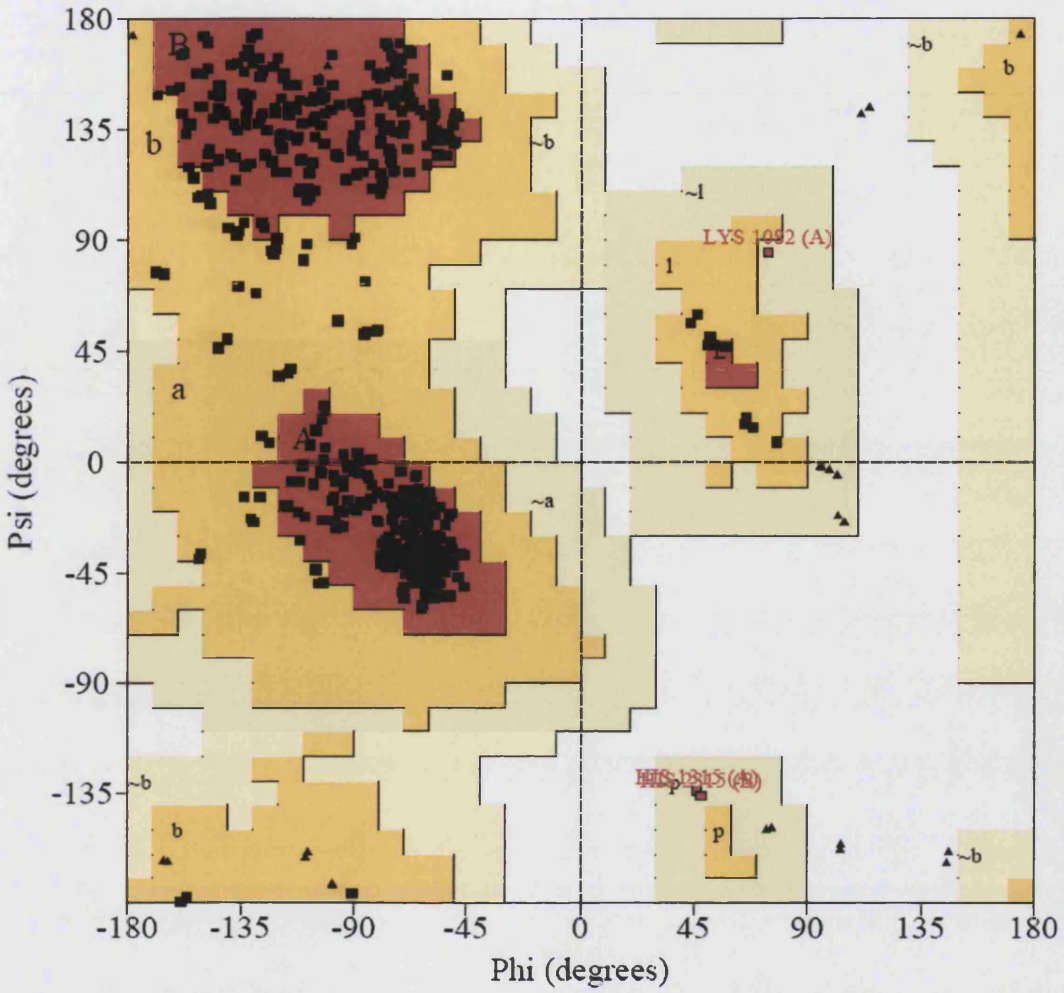


Figure 3.20: Ramachandran plot for the refined structure of Asp316Glu mutant of α3GT in complex with UDP and LacNAc at 1.76 Å resolution. The red coloured region of the plot (*A, B, L*) represents the residues in the most favoured regions; yellow regions (*a,b,l,p*) represents residues in the additional allowed regions; grey/buff coloured regions (*~a,~b,~l,~p*) represents residues in generously allowed regions and white areas of the plot represents the sterically disallowed regions. The glycine residues are represented by triangles; *A/a/~a* represent alpha helices; *B/b/~b* represent beta strands, *L,l,~l* represent left-handed helices; *p/~p* represent epsilon.

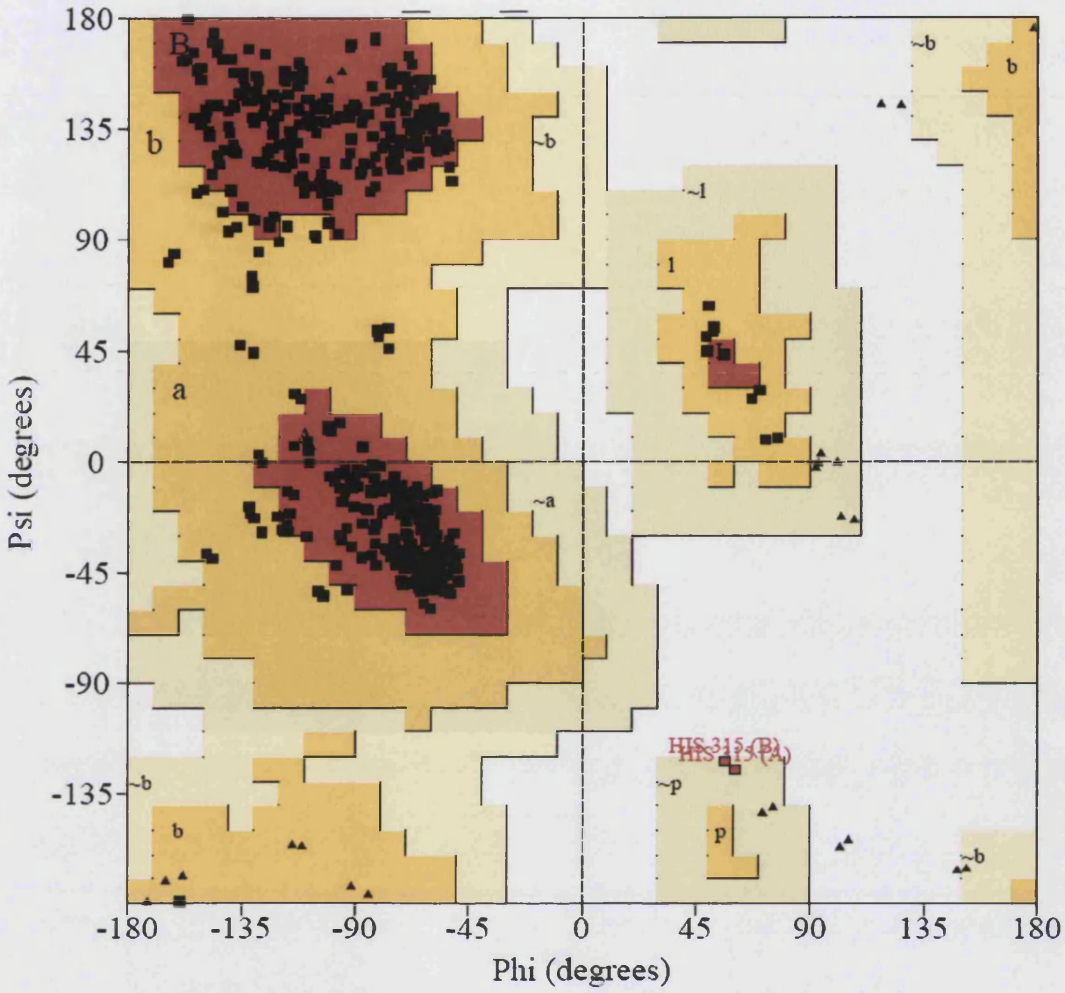


Figure 3.21: Ramachandran plot for the refined structure of Asp316Asn mutant of α3GT in complex with UDP and LacNAc at 2.2 Å resolution. The red coloured regions of the plot (*A, B, L*) represents the residues in the most favoured regions; yellow region (*a,b,l,p*) represents residues in the additional allowed regions; grey/buff coloured regions (*~a,~b,~l,~p*) represents residues in generously allowed regions and white areas of the plot represents the sterically disallowed regions. The glycine residues are represented by triangles; *A/a/~a* represent alpha helices; *B/b/~b* represent beta strands, *L,l,~l* represent left-handed helices; *p/~p* represent epsilon.

Results

Monoclinic crystals with 2 molecules per asymmetric unit were grown in the presence of UDP-Gal, manganese ion and LacNAc. Eventhough the mutants were crystallized in the presence of UDP-Gal, manganese and LacNAc, ordered electron density is only evident for the UDP moiety of UDP-Gal suggesting that either the Gal moiety was lost through hydrolysis during the experiment or the sugar moiety is disordered due to the mutation. The binding of UDP in both these structures is almost identical to the wild type enzyme in complex with UDP (Boix *et al.*, 2001). The binding of the acceptor substrate, LacNAc is essentially the same as in the wild type complex with a slight perturbation in the orientation of the GlcNAc moiety of the acceptor substrate. The area around the active site is similar to that of the wild type enzyme structure with the only difference observed at the site of the respective mutation. The orientation of Glu316 in the Asp316Glu mutant is essentially the same as the Asp316 in the wild type structure however residue Asn316 of the Asp316Asn showed a dramatic change in orientation. The position of Asn316 is almost a 180° turn from the position of Asp316 in the wild type enzyme (Figure 3.22). This might explain the extremely low activity of this mutant compared to that of Asp316Glu and Glu317Gln mutant. Mutation to Asn completely abolished the carboxylic nature of the residue, hence, compromising its stabilising interactions with UDP-Gal. The absence of any major structural change in the enzyme indicates that the reduction in catalytic activity is a direct result of the change of the residues' side chains chemistry in residue 316.

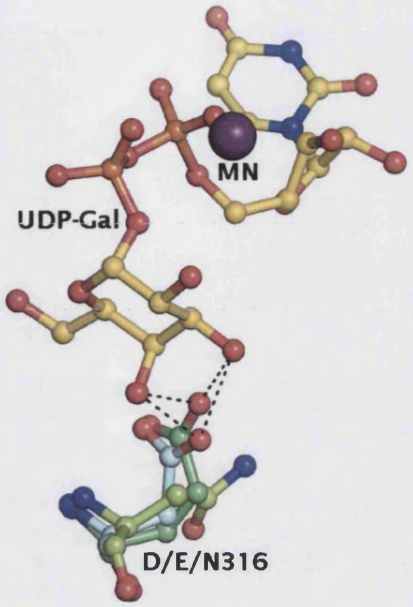


Figure 3.22: Superposition of the Glu317Gln mutant structure in complex with UDP-Gal with the Asp316Glu and Asp316Asn structures. Residue 316 in the Glu317Gln, Asp316Glu and Asp316Asn mutant structures are shown in lightcyan, limegreen and chartreuse respectively.

3.3.4 His315Arg mutant in the presence of UDP.

Methods

His315 of α3GT which is observed to interact directly with the O6' atom of galactose moiety in the Glu317Gln mutant structure corresponds to the inactivating residue in Forssman synthase (FS) enzyme. Two His315 mutant constructs to Arg which corresponds to a substitution in inactive human FS and, mutation to Gln which corresponds to a residue present in the active canine FS were created and studied (Tumbale *et al.*, data not published). Mutational studies showed that the mutation of His315 to Arg reduced the k_{cat} of the enzyme significantly whilst the mutation to Gln weakens UDP-Gal binding (Tumbale *et al.*, data not published). Crystals of His315Arg were grown in the presence of UDP-Gal were obtained although attempts to crystallise the His315Gln mutant were unsuccessful. Crystallisation of His315Arg mutant was carried out in the presence of UDP-Gal

using the same crystallisation methods described previously for the Glu317Gln and the Asp316 mutants. The dataset was scaled and processed to 2.4 Å resolution. Because the cell dimensions for the His315Arg dataset were isomorphous with the previously determined wild type structure [form II , PDB code: 1K4V (Boix *et al.*, 2001)], these coordinates were used as a starting model. Crystallographic refinement was performed using the program REFMAC (CCP4, 1994). After the initial refinement, the difference electron density maps revealed the presence of the mutated residue at position 315 and unambiguous electron density corresponding to the UDP moiety of UDP-Gal and manganese ion bound in the active site. Water molecules were gradually included into the model at positions corresponding to peaks in the $|F_o| - |F_c|$ electron density map with heights greater than 3σ and at H-bond distance from appropriate atoms. The data collection and final refinement statistics are listed in Table 3.7. Data collection and refinement statistics are shown in Table 3.7. The program PROCHECK (Laskowski *et al.*, 1993) was used to assess the quality of the final structure. Analysis of the Ramachandran plot showed all the residues for the His315Arg complex structure to be within the allowed region of the plot (Figure 3.23).

Table 3.7: Data collection and refinement statistics for His315Arg-UDP complex structure

Substrates observed in the crystal structure	Manganese
	UDP moiety
Disordered regions	none
Space group	$P2_1$
Mols/A.S.U	2
Crystal parameters	
a (Å)	44.35
b (Å)	94.23
c (Å)	94.50
α (°)	90.0
β (°)	99.08
γ (°)	90.0
Resolution (Å)	50.0-2.40
R_{symm} (%)[a] (outermost shell)[b]	9.3 (15.3)
Completeness (%) (outermost shell)[b]	90.9 (74.1)
***I/σI* (outermost shell)[b]**	7.8 (2.48)
Total reflections	69819
Unique reflections	27277
R_{cryst}[c]/R_{free}[d] (%)	18.4/21.6
Ramachandran Plot	
% core/allowed	89.5 /10.5
RMSD from ideal	
Bond angles (°)	1.113
Bond lengths (Å)	0.008
Number of water molecules	434
***B*-factor statistics (Å^2)**	
Protein[e]	25.1 (A), 27.8 (B)
Ligand in the active site[e]	21.4 (UDP-A), 22.6 (UDP-B)
Metal[e]	19.5 (MN1-A), 25.2 (MN1-B)

a $R_{symm}=\Sigma_h\Sigma_i[|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)]$, where Ii is the ith measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$.
b Outermost shell are 2.06-2.40 Å .
c $R_{cryst}= \Sigma_h|F_o - F_c|/\Sigma hF_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h.
d R_{free} is equal to R_{cryst} for a randomly selected 3.5% subset of reflections, not used in refinement.
e Temperature factors for individual molecules quoted

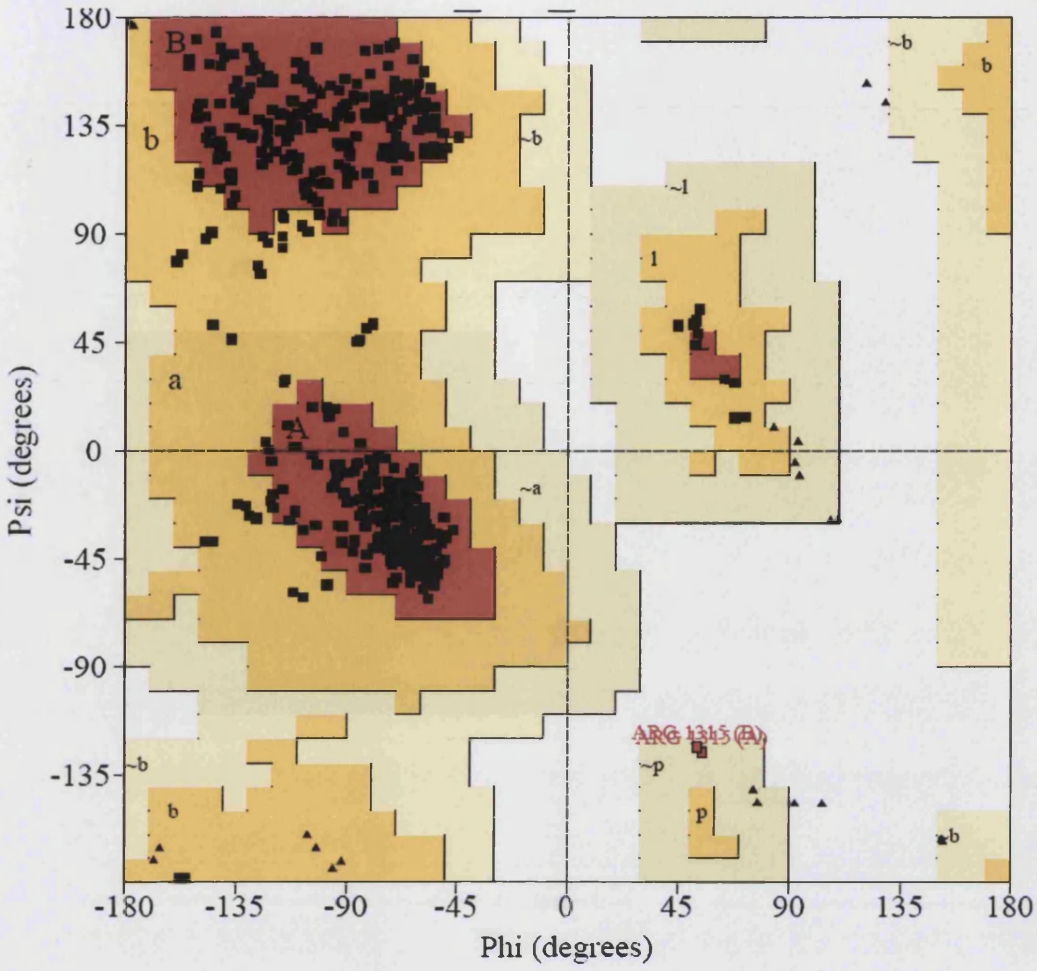


Figure 3.23: Ramachandran plot for the refined structure of the His315Arg mutant of α3GT in complex with UDP and manganese ion at 2.4 Å resolution. The red coloured region of the plot (*A, B, L*) represents the residues in the most favoured regions; yellow regions (*a,b,l,p*) represents residues in the additional allowed regions; grey/buff coloured regions (*~a,~b,~l,~p*) represents residues in generously allowed regions and white areas of the plot represents the sterically disallowed regions. The glycine residues are represented by triangles; *A/a/~a* represent alpha helices; *B/b/~b* represent beta strands, *L,l,~l* represent left-handed helices; *p/~p* represent epsilon.

Results

The crystal structure of the His315Arg mutant showed binding of only the UDP moiety despite the fact that crystallisation was carried out in the presence of UDP-Gal, indicating that hydrolysis of UDP had occurred. Crystallisation was carried out in the absence of acceptor substrate however, galactose can still be transferred to water albeit at a much slower rate. The structure of this mutant is exactly the same as previous structures of the wild type mutant and the Asp316 mutant complexes indicating that the crystallisation process is promoted after cleavage of UDP-Gal and that the formation of the dead-end UDP•enzyme complex optimises crystal formation in the $P2_1$ space group. The position of the side chain of mutant residue Arg315 is approximately 5 Å away from the bound UDP however superposition of the Glu317Gln•UDP-Gal structure with the His315Arg structure showed that the longer side chain of Arginine might interfere with the binding of UDP-Gal or at the very least hinder the stabilisation of the folded conformation of UDP-Gal (Figure 3.24).

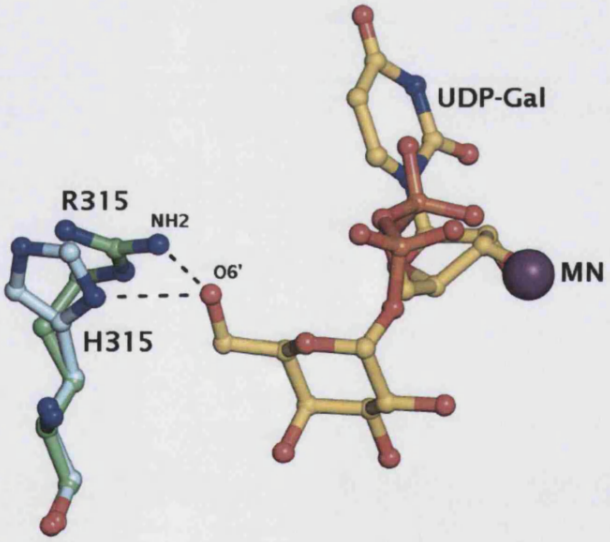


Figure 3.24: Superposition of UDP-Gal from the Glu317Gln mutant structure with the His315Arg mutant structure shows that the O6' of the galactose moiety would be too close to the NH2 atom of the mutated Arg315 residue. Residues His315 from the Glu317Gln structure and Arg315 in the His315Arg mutant are shown as palecyan and palegreen sticks respectively while UDP-Gal is shown as yellow sticks. The close interactions between the 315 residues and UDP-Gal are shown as black dashed lines.

3.3.5 Discussion

It has been proposed that the catalytic reaction for retaining glycosytransferases proceeds through a mechanism involving a covalently bound glycosyl-enzyme intermediate (Breton *et al.*, 2006; Lairson and Withers, 2004; Unligil and Rini, 2000). If this scheme is to be followed, a suitable nucleophile should be located close to the anomeric carbon (C1) of the UDP-Gal and on the correct (β) face to allow the displacement of the UDP leaving group. Previous structural studies on α3GT had identified residue Glu317 as a possible candidate for this role (Gastinel *et al.*, 2001) however subsequent mutational and structural studies failed to support this theory (Boix *et al.*, 2002; Boix *et al.*, 2001; Zhang *et al.*, 2003), hence casting the role of Glu317 as a catalytic nucleophile in the proposed double displacement mechanism in doubt. In recent years, a number of crystal structures representing several families of retaining transferases have been solved however due to the high hydrolase activity, only a fraction of these structures have intact donor substrates bound in the active site hence limiting the identification of possible catalytic nucleophiles for modeling attempts. Therefore, the determination of crystal structures with uncleaved donor substrates in the active site is highly anticipated. The crystal structure of the Glu317Gln mutant in the presence of the active substrate, UDP-Gal showed the intact substrate bound in the active site, providing insight into the interactions which occur between the active site residues and the donor substrate prior to hydrolysis of the sugar. The distorted conformation of the UDP-sugar is reflective of donor-enzyme complex structures of other glycosyltransferases and this suggests that this distorted conformation might play a role in the catalytic mechanism of glycosyltransferases. However this structure does not provide great support for the double displacement mechanism but instead tends to strengthen the S_Ni type mechanism hypothesis which was previously proposed based on modeling studies (Boix *et al.*, 2002). In this mechanism, it is proposed that O3 of the acceptor displaces the UDP on the same side of the galactose ring to produce an α-galactosyl product without formation of a β-galactosyl-enzyme intermediate. At the same time, O2 of the β-phosphate of the UDP is positioned to deprotonate the acceptor O3

deprotonate the acceptor O3 (Boix *et al.*, 2002).Although we currently do not have a crystal structure of the enzyme in complex with both the intact donor and acceptor substrate, superposition of the Glu317Gln-UDP-Gal complex structure with that of the wild type enzyme in complex with UDP and lactose (PDB code: 1GX4) suggests that this might be a plausible mechanism for α3GT. The hydroxyl O3 atom from the galactose moiety of N-acetyllactosamine is 2.5 Å away from the anomeric carbon C1' of UDP-Gal and the lack of any other polar atoms within 4 Å of the reactive center C1' suggests that the acceptor substrate might be a viable candidate for a nucleophile (Figure 3.31).

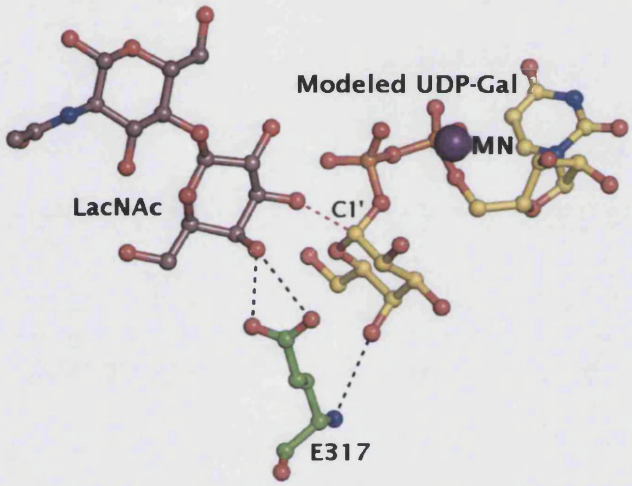


Figure 3.31: Structure of the wild type α3GT in complex with GalNAc (PDB code: 1GX4(Boix *et al.*, 2002)) modeled with UDP-Gal from the Glu317Gln structure. The red dashed line shows the distance between LacNAc and C1' of UDP-Gal which is approximately 2.53 Å.

A water molecule which forms direct hydrogen bonds with O6' and O5' of the galactose moiety as well as O1B on the β-phosphate of UDP is also in close proximity to the C1' carbon (3.87 Å). The critical residue, 317 is positioned on the β-face of the galactose moiety however, it is too far from the C1' carbon (4.8 Å) to effectively act as a nucleophile in support of the double displacement mechanism but on the other hand, the possibility of structural changes occurring to allow a close

approach of residue 317 to C1' in the ternary enzyme complex cannot be totally precluded. The kinetic data showed that the mutation to Gln does cause a significant decrease in the catalytic activity of the enzyme however, the magnitude of this change does not lend support to the proposal that Glu317 is the catalytic nucleophile but instead suggests that Glu317 plays a role in acceptor substrate binding and in stabilizing a cationic transition state for galactose hydrolysis (Zhang *et al.*, 2003). The structure of the Glu317Gln mutant in complex with UDP-Gal also showed disorder of the C-terminus with the exception of residue 365 which shows obvious side chain density for the Arg residue. The Arg side chain is shown to interact directly with the O2A atom of the α-phosphate of UDP in a similar way as in the closed C-terminus structure (Form II) indicating that this residue might play a role in catalysis. The disorder of the C-terminus in the absence of acceptor probably indicates that binding of the acceptor and cleavage of the UDP-Gal bond might occur concurrently with the structural change of the C-terminus to the conformation which folds over to form the full active site. This structural change leads to the formation of the active site pocket which serves to provide an ideal environment for the binding of the acceptor molecule in the correct orientation for catalysis. The occlusive nature of the active site pocket facilitates efficient catalysis by ensuring that water molecules are excluded from the active site, or at least controlled within it. The bent conformation of the bound donor, UDP-Gal besides ensuring optimal binding by minimizing steric hindrance also acts to present the C1' carbon for nucleophilic attack by the approaching acceptor molecule. It is possible that upon binding of the acceptor substrate the galactose ring of UDP-Gal is bent further underneath the diphosphate plane of UDP making the C1' carbon more vulnerable to nucleophilic attack from O13 atom of the acceptor.

Residue Asp316 is predicted to play a role in stabilizing and controlling the folded conformation of the galactose ring in order for hydrolysis to occur. The kinetic data shows that mutation of this residue to glutamate decreased the k_{cat} to 31% of the wild type activity and this decrease in catalytic activity maybe attributed to the change in side change structure. Although we only managed to obtain the structure of Asp316Glu mutant in complex with a hydrolysed UDP-Gal (UDP) and

N-acetyllactosamine, when this structure is superimposed with the structure of the Glu317Gln in complex with UDP-Gal it is obvious that the longer side chain of Glu would have made a much closer and stronger hydrogen bonding interaction with the O4' atom of galactose. This might have compromised the ability of the sugar moiety to be transferred to the acceptor molecule, furthermore, the longer side chain may have served as a barrier for the galactose ring to fold further underneath the diphosphate. Mutation of residue Asp316 to Asn renders the protein inactive and this could be due to the change in side chain chemistry which could have compromised the hydrogen bond interaction with the O4' atom of galactose. Modelling of UDP-Gal in the crystal structure of the Asp316Asn mutant showed that the Asn residue faced away from the galactose moiety and would not have made any direct interactions with the galactose. Asp316 is expected to function in stabilizing the conformation of the sugar in the correct orientation for hydrolytic attack by the acceptor molecule.

3.4 Structural basis for substrate specificity of α3GT

3.4.1 Structure of AGGL (His280Ala-Ala281Gly-Ala282Gly-Ile283Leu) mutant in the presence of UDP-Gal

Introduction

His280 has been predicted to be responsible for conferring substrate specificity in α3GT (Gastinel *et al.*, 2001). Several mutant constructs with substitutions of Gln, Asn, Thr, Arg, Ala, Gly, Leu, and Pro for His280 based on the variability of the residue corresponding to His280 in different relatives of α3GT were generated by Zhang and co-workers and studied (Zhang *et al.*, 2003). These mutants were shown to have very low enzymatic activity with only the mutation to Gln displaying activity greater than 1% of the wild-type enzyme (Zhang *et al.*, 2003). Attempts to co-crystallise the His280 mutants with donor UDP-Gal were carried out in the hope of obtaining a structure of the mutant in complex with the un-cleaved donor substrate and to shed light on how His280 can structurally confer substrate

specificity of α3GT towards UDP-Gal. All of the His280 mutants were less soluble than the wild type enzyme and tend to precipitate on addition of the donor substrate, UDP-Gal. This made characterisation and crystallisation of the mutants in the presence of substrates almost impossible. Due to this factor, the AGGL mutant was constructed based on the structure of the Glu317Gln•UDP-Gal complex in order to study the donor substrate specificity of the enzyme. In the Glu317Gln complex structure, it is observed that the O2' atom of the galactose moiety interacts directly with His280 which has been previously implicated in the substrate recognition and specificity of α3GT. Previous mutational studies of His280 showed that mutations to smaller side chains, Alanine and Glycine showed low levels of GalNAc transferase activity while mutation to a larger side chain, Arginine rendered the enzyme completely inactive (Zhang *et al.*, 2003) which contradicted previous studies on porcine α3GT by Lazarus and co workers which stated that mutation of His271 (porcine equivalent to bovine His280), a residue critical for donor substrate recognition to smaller side chain residues such as Ala, Leu and Gly caused complete loss of activity whilst mutation to Arg retained the enzyme's activity. (Lazarus *et al.*, 2002). Characterisation of the mutant showed that the donor substrate specificity was altered from UDP-Gal to UDP-GalNAc. It displayed a marked increase in UDP-GalNAc hydrolysis and a lower activity for Gal transfer (Tumbale *et. al*, unpublished results).

Methods

Crystallisation of this mutant in the presence of UDP-GalNAc was sought in order to observe the interactions of the GalNAc moiety with residues of the AGGL mutant and shed light into how changes in the side chain chemistry of the respective residues affect substrate binding and specificity. This mutant was relatively easy to crystallise compared to the other His280 mutants described previously. Crystals of the AGGL mutant grew in the presence of UDP-GalNAc and in conditions containing 10-15% PEG 6000, 15-25% MPD and 0.1M Tris-HCl, pH 8. No additional cryoprotectant was used during data collection as the high concentration of MPD in the crystallisation condition could sufficiently serve as cryoprotectant. A

dataset was collected, processed and scaled to 2.7 Å in $P3_121$ space group with one molecule in the asymmetric unit. The structure was solved by the molecular replacement method using the native form II structure (PDB code: 1K4V (Boix *et al.*, 2001) as a search model with the program Molrep (Vagin and Teplyakov, 1997). Crystallographic refinement was performed using the program package REFMAC (CCP4, 1994). Residues with poor side-chain density were modelled as alanines and regions with very poor main chain electron density were excluded from refinement as not to bias the model. Water molecules were gradually included into the model at positions corresponding to peaks in the $|F_o| - |F_c|$ electron density map with heights greater than 3σ and at H-bond distance from appropriate atoms. Refinement was carried out until the R_{cryst} and R_{free} could not be improved. The data collection statistics and the final refinement statistics are listed in Table 8. The program PROCHECK (Laskowski *et al.*, 1993) was used to assess the quality of the final structure. Analysis of the Ramachandran plot showed all the residues for the His280Arg structure to be within allowed region of the plot (Figure 3.26).

Table 3.9: Data collection and refinement statistics for the AGGL mutant structure

Substrates observed in the crystal structure	Manganese UDP moiety
Disordered regions	340-368, 193-197
Space group	$P3_121$
Mols/A.S.U	1
Crystal parameters	
a (Å)	70.2
b (Å)	70.2
c (Å)	127.0
α (°)	90.0
β (°)	90.0
γ (°)	120.0
Resolution (Å)	50.0-2.7
R_{symm} (%)[a] (outermost shell)[b]	13.7 (30.3)
Completeness (%) (outermost shell)[b]	99.7 (98.3)
***I/σI* (outermost shell)[b]**	12.1 (4.2)
Total reflections	89132
Unique reflections	11624
R_{cryst}[c]/R_{free}[d] (%)	26.0/19.2
Ramachandran Plot	
% core/allowed	88.6/11.4
RMSD from ideal	
Bond angles (°)	0.010
Bond lengths (Å)	1.199
Number of water molecules	44
***B*-factor statistics (Å^2)**	
Protein[e]	24.4
Ligand in the active site[e]	20.6 (UDP)
Metal[e]	20.8 (MN1)

a $R_{symm}=\Sigma_h\Sigma_i[|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)]$, where Ii is the ith measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$.
b Outermost shell are 2.77-2.7 Å.
c $R_{cryst}= \Sigma_h|F_o - F_c|/\Sigma hF_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h.
d R_{free} is equal to R_{cryst} for a randomly selected 8% subset of reflections, not used in refinement.
e Temperature factors for individual molecules quoted

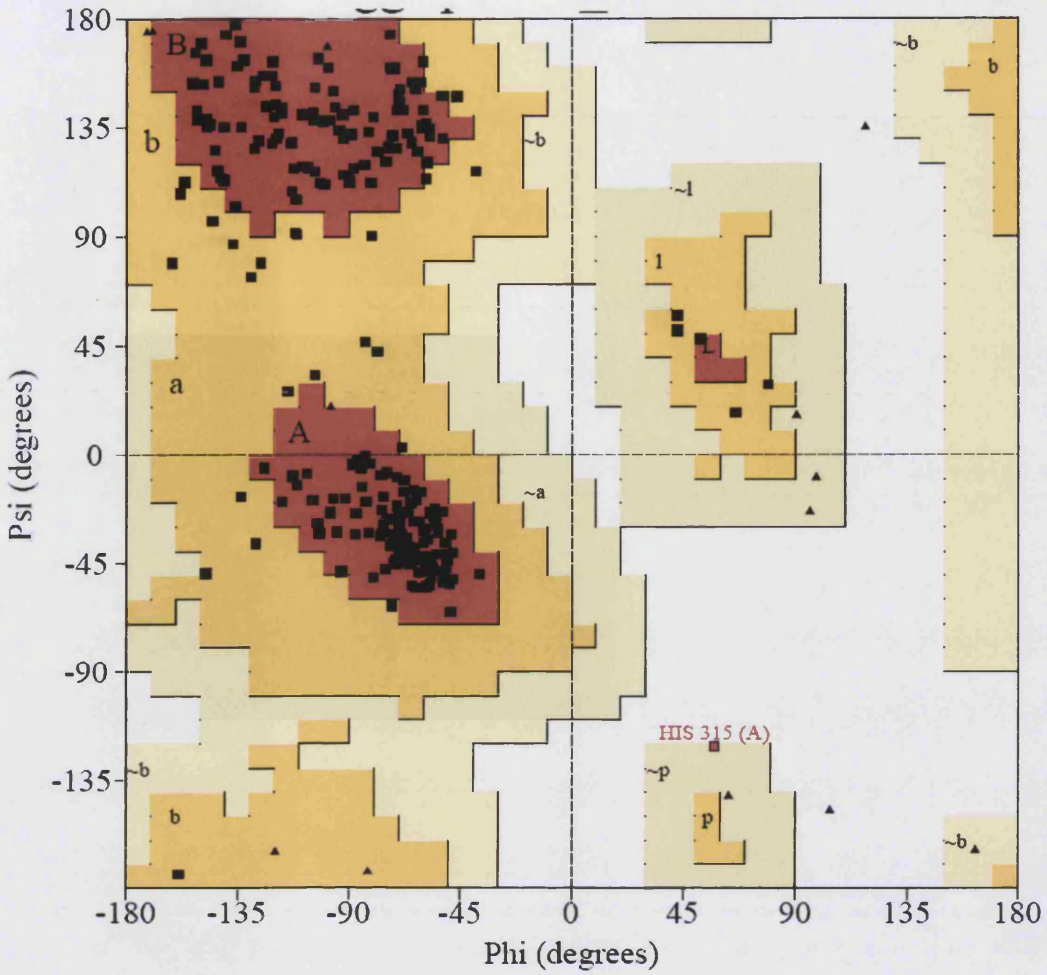


Figure 3.26: Ramachandran plot for the refined structure of the AGGL mutant of α3GT in complex with UDP at 2.7 Å resolution. The red coloured region of the plot (*A, B, L*) represents the residues in the most favoured regions; yellow regions (*a,b,l,p*) represents residues in the additional allowed regions; grey/buff coloured regions (*~a,~b,~l,~p*) represents residues in generously allowed regions and white areas of the plot represents the sterically disallowed regions. The glycine residues are represented by triangles; *A/a/~a* represent alpha helices; *B/b/~b* represent beta strands, *L,l,~l* represent left-handed helices; *p/~p* represent epsilon.

Results

The AGGL mutant crystallised in the trigonal space group with one molecule per asymmetric unit. The crystal structure of the mutant showed binding of the manganese ion and UDP without the GalNAc moiety even though it was crystallised with UDP-GalNAc indicating that hydrolysis might have take place. The mutations for the respective residues His280 to Ile283 were also observed in the structure clearly. The mutations left a pocket in the donor substrate recognition site that would accommodate the bulky 2-acetamido group of GalNAc (Figure 3.27) hence, explaining the relatively higher activity that this mutant has towards GalNAc as compared to Gal (Tumbale *et al.*, unpublished data). This indicates that the mutation corresponding to residues 280-283 to smaller residues altered the specificity of the enzyme for UDP-GalNAc.

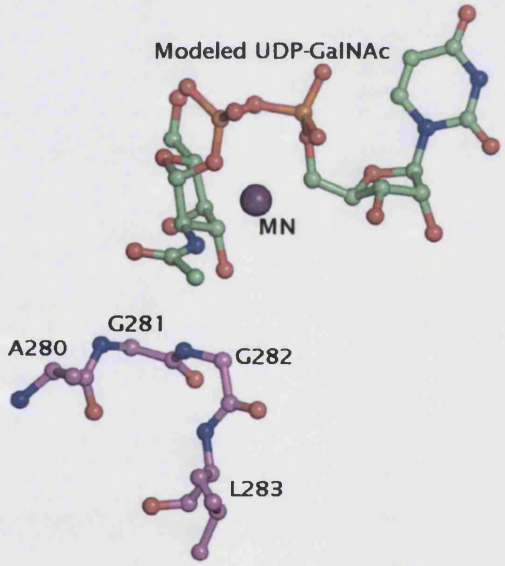


Figure 3.27: AGGL mutant structure modeled with UDP-GalNAc. Residues from the AGGL mutant are shown in violet while the modeled UDP-GalNAc is shown in limegree

3.4.2 Structure of the 5 site (His280Gly-Cys338Leu-Gln247His-Ala248Pro-Trp249Gly) mutant in the presence of donor substrates

Methods

The 5 site mutant was constructed in order to study the effects of mutating residues identified as important for donor and acceptor binding on the catalytic efficiency of the enzyme. Crystallisation of the 5 site mutant in the presence of both donor and acceptor substrates was carried out in the hope of obtaining a crystal structure of α3GT with both the intact UDP-Gal and LacNAc bound in the active site. Crystals of the 5 site mutant were grown in the presence of 10 mM $MnCl_2$, 10mM UDP and UDP-gal respectively in crystallisation condition containing 10-15% PEG 4000, 0.1 M Tris-HCl, pH 8.0 and 0.2 M sodium acetate. Before data collection the crystals were flash-cooled at 100K in a cryoprotectant containing the reservoir solution and 25% glycerol. 5 site mutant datasets to 2.1 Å and 1.88 Å resolution respectively (station PX 14.2, SRS Daresbury, using an Area Detector Systems Corp. detector system) were collected. Raw data images were indexed and scaled using the DENZO and SCALEPACK modules of the HKL suite (Otwinowski and Minor, 1997). The structures of the 5 site mutants were determined by the molecular replacement method using the native form II structure [PDB code: 1K4V (Boix *et al.*, 2001)], as a search model with the program MOLREP (Vagin and Teplyakov, 1997). Crystallographic refinement was performed using the program REFMAC. After the initial refinement, the difference electron density maps revealed the presence of all 5 mutated residues at their respective positions. Water molecules were gradually included into the model at positions corresponding to peaks in the $|F_o| - |F_c|$ electron density map with heights greater than 3σ and at H-bond distance from appropriate atoms. Residues with poor side-chain density were modelled as alanines and regions with very poor main chain electron density were excluded from refinement so as not to bias the model. Refinement was carried out until the R_{cryst} and R_{free} could not be improved. The data collection statistics and the final refinement statistics are listed in Table 3.10. The program PROCHECK (Laskowski *et al.*, 1993) was used to assess the quality of the final structure. Analysis of the

Ramachandran plot showed all the residues for the 5 site mutant structures to be within the allowed regions of the plot (Figure 3.28 and 3.29 respectively).

Table 3.10: Data collection and refinement statistics for the 5 site mutant grown in the presence of UDP and UDP-Gal respectively

	5 site mutant (His280Gly-Cys338Leu-Gln247His-Ala248Pro-Trp249Gly)	
Substrates used in crystallisation	10mM UDP 10mM Manganese	10mM UDP-Gal 10mM Manganese
Substrates observed in the crystal structure	none	none
Disordered regions	191-198 342-368	343-368
Space group	P32	P32
Mols/A.S.U	2	2
Crystal parameters		
a (Å)	66.79	64.94
b (Å)	66.79	64.94
c (Å)	136.73	136.53
α (°)	90.0	90.0
β (°)	90.0	90.0
γ (°)	120.0	120.0
Resolution (Å)	50.0-2.1	50-1.88
R_{symm} (%)[a] (outermost shell)[b]	6.4 (34.9)	7.3 (46.0)
Completeness (%) (outermost shell)[b]	98.0 (80.7)	95.6 (87.5)
***I/σI* (outermost shell)[b]**	28.33 (3.79)	17.19 (2.66)
Total reflections	565302	609673
Unique reflections	39902	52337
R_{cryst}[c]/R_{free}[d] (%)	22.2/25.19	19.9/24.4
Ramachandran Plot % core/allowed	88.7/11.3	92.0/8.0
RMSD from ideal		
Bond angles (°)	1.256	1.132
Bond lengths (Å)	0.009	0.009
Number of water molecules	328	408
***B*-factor statistics (Å^2)**		
Protein[e]	27.2 (A), 27.3 (B)	25.0 (A), 25.0 (B)
Ligand in the active site[e]	-	-
Metal[e]	-	-

a $R_{symm}=\Sigma_h\Sigma_i[|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)]$, where I_i is the ith measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of $I(h)$.
b Outermost shell are 2.16-2.11 and 1.93-1.88 respectively .
c $R_{cryst}= \Sigma_h|F_o - F_c|/\Sigma hF_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h.
d R_{free} is equal to R_{cryst} for a randomly selected are 2.8% and 2.1% subset of reflections respectively, not used in refinement.
e Temperature factors for individual molecules quoted

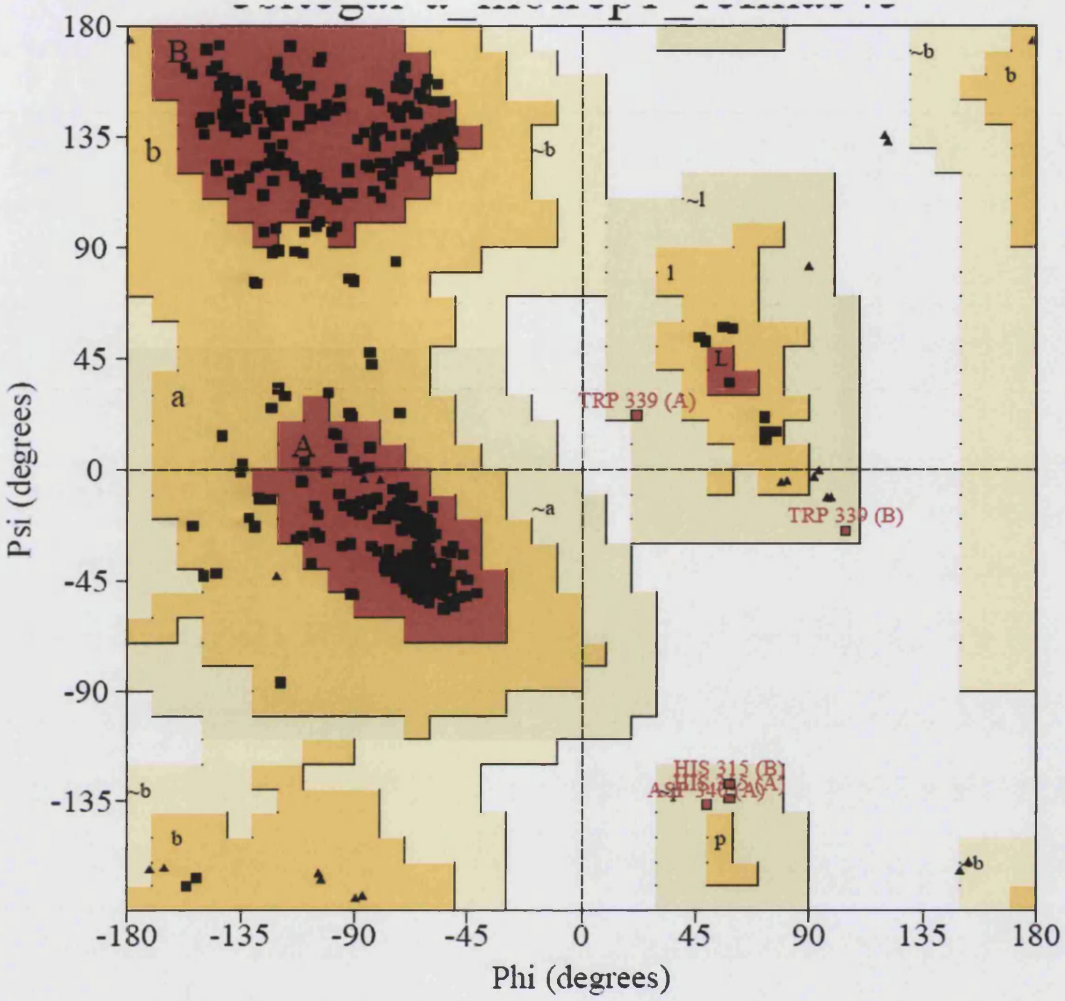


Figure 3.28: Ramachandran plot for the refined structure of the 5 site mutant of α3GT in the presence of UDP at 2.1 Å resolution. The red coloured regions of the plot (*A, B, L*) represents the residues in the most favoured regions; yellow regions (*a,b,l,p*) represents residues in the additional allowed regions; grey/buff coloured regions (*~a,~b,~l,~p*) represents residues in generously allowed region and white areas of the plot represents the sterically disallowed regions. The glycine residues are represented by triangles; *A/a/~a* represent alpha helices; *B/b/~b* represent beta strands, *L,l,~l* represent left-handed helices; *p/~p* represent epsilon.

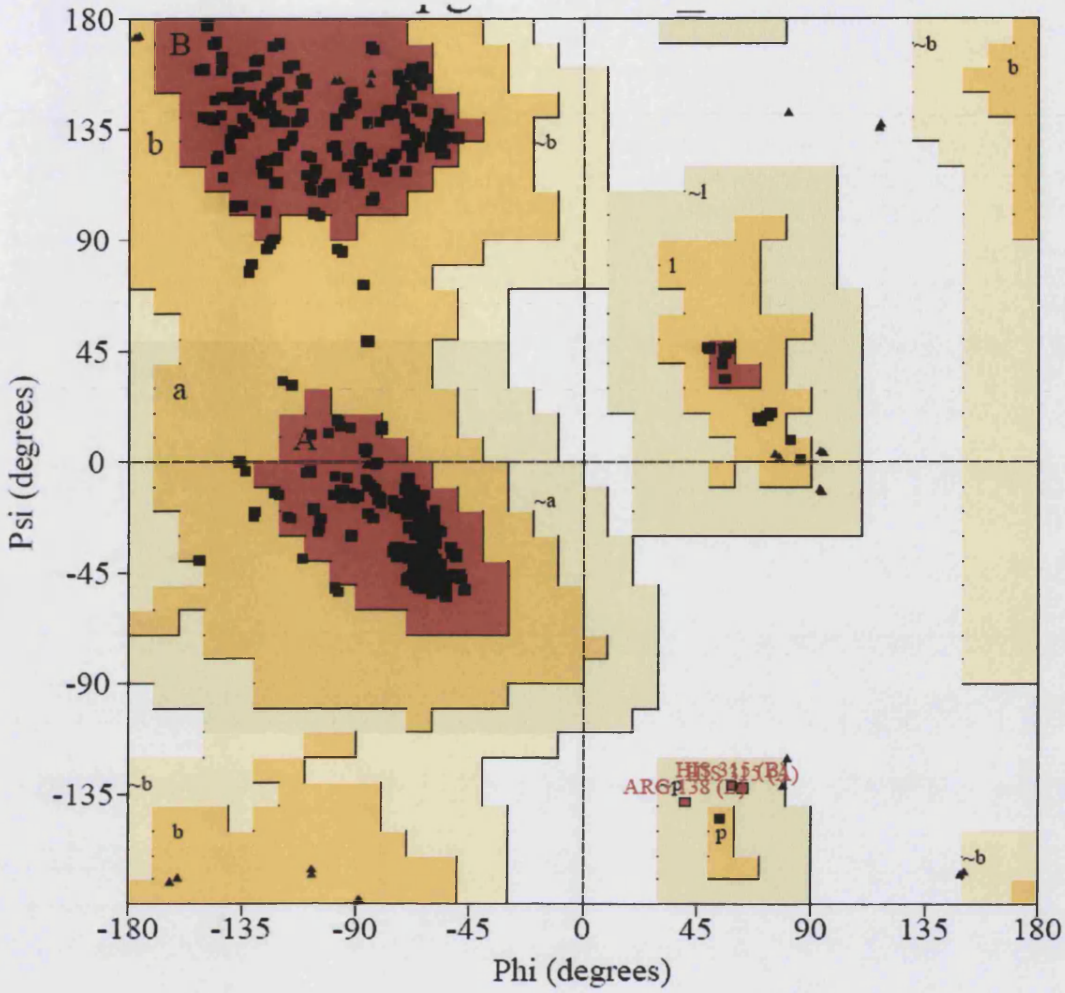


Figure 3.29: Ramachandran plot for the refined structure of the 5 site mutant of α3GT in the presence of UDP-Gal at 1.88 Å resolution. The red coloured regions of the plot (*A, B, L*) represents the residues in the most favoured regions; yellow region (*a,b,l,p*) represents residues in the additional allowed regions; grey/buff coloured regions (*~a,~b,~l,~p*) represents residues in generously allowed regions and white areas of the plot represents the sterically disallowed regions. The glycine residues are represented by triangles; *A/a/~a* represent alpha helices; *B/b/~b* represent beta strands, *L,l,~l* represent left-handed helices; *p/~p* represent epsilon.

Results

The crystal structures of the 5 site mutant in the presence of UDP and UDP-Gal are in the trigonal space group with 2 molecules in the asymmetric unit. This structure did not show any ligand binding even though the crystallisation was carried out in the presence of donor substrates and manganese ions. Interestingly this structure showed significant amounts of disorder and rearrangement in several regions of the enzyme. The disorder and rearrangement is thought to be mostly attributed to the respective mutations which had been carried out. Although electron density for most of the polypeptide chain of this mutant is very well defined, there are several regions which exhibit very poor electron density. Exclusion of these regions from refinement showed almost complete absence of electron density for region 192-198 and region 342-368 (Figure 3.30A) for the 5 site mutant structure in the presence of UDP whilst the 5 site mutant in the presence of UDP-Gal only showed disorder in region between residues 343-368. The disorder for region 192-198 is similar to the structure of the 'apo' wild type enzyme in previous studies (Jamaluddin *et al.*, 2007). This seems to suggest that the 192-198 region is highly flexible and undergoes a conformational change on binding of the donor substrate (Jamaluddin *et al.*, 2007). The rearrangement and disorder in the region 342-368 is observed to be attributed to the mutation of residue Trp249 to glycine. In the wild type enzyme structure Trp249 is observed to make a direct interaction with Asp340 (Figure 3.30B). We believe that this interaction is important in maintaining the structural integrity of the enzyme, therefore mutation to glycine in the 5 site mutant would disrupt this crucial interaction hence destabilising the region corresponding to residues 342 to residue 368. The large structural changes that are observed in this mutant may be one of the reasons why this mutant displayed very low activity and why we fail to observe binding for the donor substrate despite being crystallised in the presence of UDP and UDP-Gal respectively. Interestingly the disorder and rearrangement observed in these structures are similar to that of the His280Arg mutant structure.

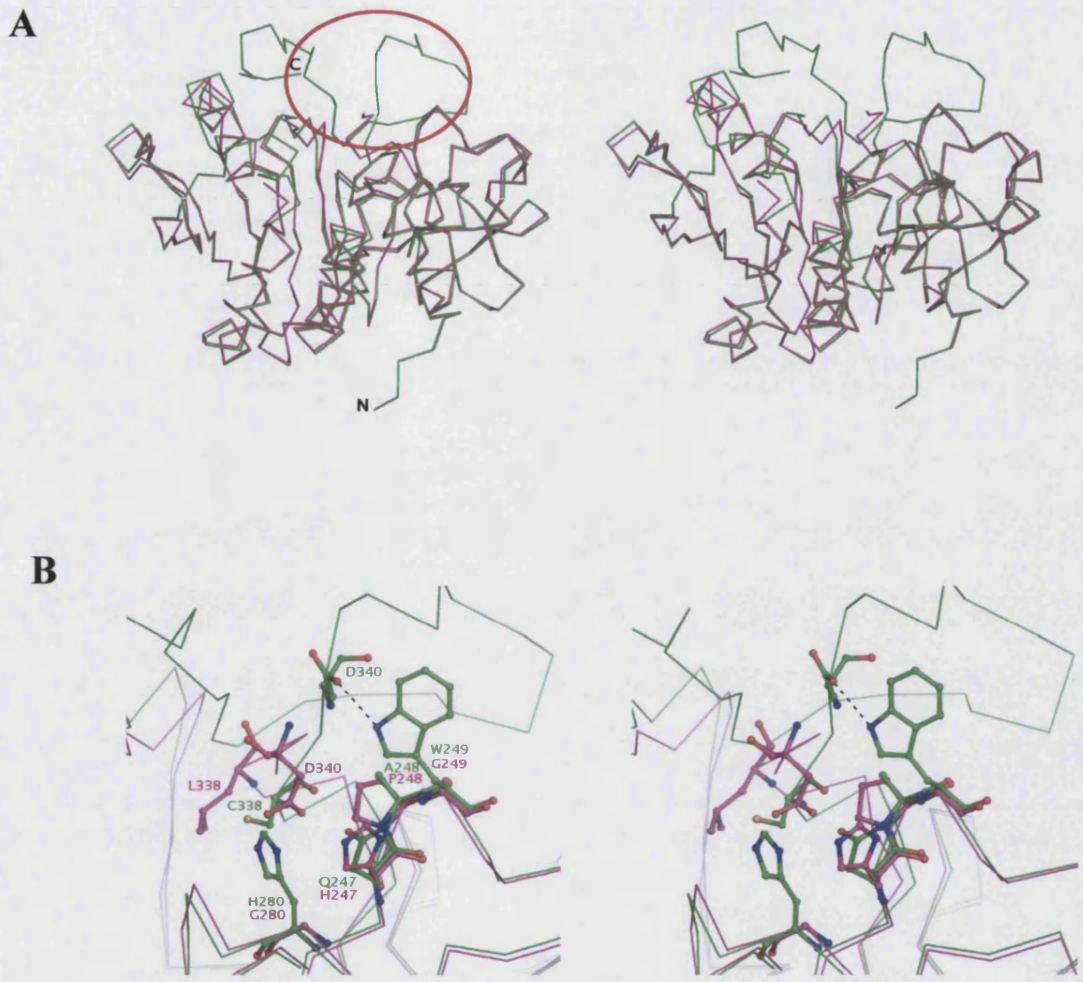


Figure 3.30- A: Cα trace of the wild type α3GT structure (green) with the structure of the 5-site mutant in the presence of UDP (magenta). The region highlighted by a red circle shows the region of the mutation. B: 5 mutated residues and the effect of the mutation has on the structural integrity of the enzyme. Mutation of Trp249 to glycine disrupted the interaction with Asp340 hence causing disorder in the region beyond Asp340.

3.4.3 Discussion

The AGGL mutant showed remarkably low galactosyltransferase activity but increased GalNAc transferase activity indicating that the quadruple mutation has altered the specificity of the enzyme. Although we could not obtain a structure of the AGGL mutant in complex with an intact UDP-GalNAc, modeling of GalNac in the active site of the mutant suggests that there is enough space to accommodate the 2-acetamido group of GalNAc at the substrate recognition pocket. Changes in donor substrate specificity due to mutagenesis have been reported for other glycosyltransferases with the most recent being of the human blood group B galactosyltransferase in which mutation of a residue adjacent to the metal binding motif changed the specificity of the enzyme from UDP-Gal to UDP-GalNAc (Persson *et al.*, 2007). Other reported studies include mutants of bovine 4-β-galactosyltransferase which displayed increased glucosyltransferase activity upon mutation of a single residue, Arg228Lys (Ramakrishnan *et al.*, 2005) and increased GalNAc transferase activity with a Tyr289Leu mutation (Ramakrishnan and Qasba, 2002). The donor specificity of toxin B glucosyltransferase from *C. difficile* altered to UDP-GalNAc upon a double mutation (Jank *et al.*, 2005). A plant anthocyanin galactosyltransferase altered its specificity to glucosyltransferase with a Gln382His mutation (Kubo *et al.*, 2004). This shows how altering the size and shape of the substrate recognition pocket in glycosyltransferase via mutation can change the specificity of the enzyme hence, offering a promising scheme for structure based protein engineering to generate glycosyltransferases of desired specificity.

The 5 site mutant has mutations on residue His280 and Trp249 indicating that the disorder which is observed in both structures might be due to the structural change of the side chain residue. In the wild type enzyme structure, His280 is observed to interact with residue Trp356 via a water molecule and residue Trp249 is observed to interact directly with Asp340. In the 5 site mutant, His280 and Trp249 are mutated to glycines, totally eliminating the side chain structures that coordinate these interactions and in effect compromising the structural integrity of the enzyme hence, causing rearrangement and disorder of the region beyond residue 337. This

result is indicative of how mutations can compromise the structural integrity of a protein.

3.5 Conclusions

The results presented here serve to highlight the role of the residues lining the active site in particular residue Glu317 which previously has been implicated as the catalytic nucleophile for a proposed double displacement mechanism. Structural and mutational studies that we have carried out however, points to a possibility of a different mode of mechanism in which the acceptor molecule itself acts as the catalytic nucleophile. We predict that Glu317 acts to facilitate acceptor substrate binding and stabilize a cationic transition state for galactose cleavage. The formation of the ternary complex for catalysis is inherently linked to the conformational change of the C-terminus which provides an enclosed environment for efficient catalysis by minimizing unwanted hydrolysis. His280 plays an important role in catalysis by conferring the donor substrate specificity of the enzyme. There is potential for structure based rational design of α3GT with differing donor substrate specificity.

3.4.4 Crystal structure of α3GT in complex with *p*-Nitrophenyl galactoside (pNPGal) and modeling studies of α3GT with the inhibitor U66.

Introduction

Enzymes are nature's very own catalyst and they exhibit unparalleled catalytic prowess and environmental compatibility. Despite these attractive attributes enzymes have yet to witness widespread application in synthetic chemistry. The under use of their enormous potential stems not only from the reluctance of synthetic chemists to deviate from their conventional chemical methods but also from the general limitations on enzyme applicability that arise from the high degree of substrate specificity displayed by most enzymes. However, concerted efforts by means of rational protein engineering and directed evolution methods to alter substrate specificity are successfully challenging the perceived latter limitation.

Another alternative strategy for the expansion of an enzyme's substrate specificity hence, its synthetic application is through 'substrate engineering'. The attachment of a readily removable functional group to an alternative glycosyltransferase substrate induces a productive binding mode, facilitating rational control of substrate specificity and regioselectivity using wild-type enzymes (Lairson *et al.*, 2006). Furthermore recent studies in the use of glycosyltransferase to synthesise glycoconjugates for therapeutic use has sparked interest in the application of alternative donor and acceptor substrates for the assembly of variable forms of glycoconjugates. Here we present the crystal structure of the wild type α3GT in complex with UDP and an alternative acceptor molecule *p*-Nitrophenyl galactoside (pNPGal). This chapter will also discuss modeling studies of an inhibitor of α3GT which is designated as U66. This chapter will also move on to describe soaking and co-crystallisations attempts which have been carried out in order to obtain the crystal structure of α3GT in complex with the inhibitor U66.

3.4.4.1 Crystal structure of wild type α3GT in complex with *p*-Nitrophenyl galactoside (pNPG).

Methods

Preparation of the wild type α3GT sample was carried out as previously described. Wild type complex crystals were grown in 5% PEG 6000 and 0.1M Tris-HCl, pH 8 in the presence of 10mM $MnCl_2$ and 10mM UDP. The wild type crystals (Figure 3.32) were then subjected to soaks in total of 50 mM pNPGal with gradual increment in concentration of 10 mM over a period of 5 days prior to X-ray exposure.

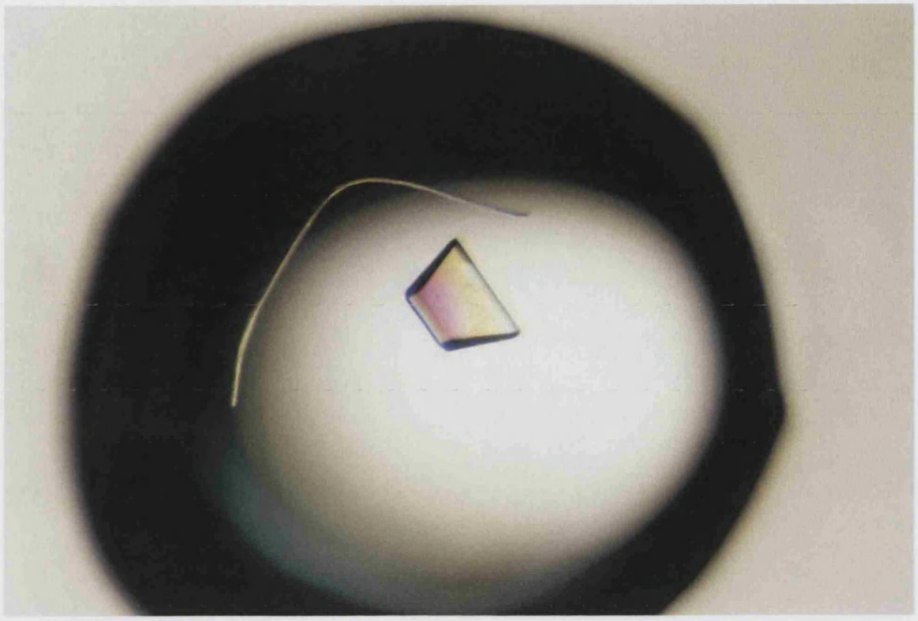


Figure 3.32: Typical crystal of wild type α3GT in a 2μl drop grown in 5-10% PEG 6000 and 0.1M Tris-HCl, pH 8; in the presence of 10mM $MnCl_2$ and 10mM UDP. These crystals will usually diffract to better than 2 Å resolution (maximum crystal size: 0.7 X 0.4 X 0.3 mm)

As the cell dimensions were isomorphous with the wild type structure, the published structure of the form II wild type enzyme [PDB code: 1K4V (Boix *et al.*, 2001)] was used directly in refinement. Crystallographic refinement was performed using the program REFMAC (CCP4, 1994). After the initial refinement, the difference electron density maps revealed unambiguous density corresponding to pNPGal bound at the acceptor binding site of the enzyme. Defined and ordered electron density is also observed for the UDP molecule and the manganese ion. Several rounds of refinement and model building using the program COOT (Emsley, 2004) were performed until the R_{free} could not be improved. Water molecules were gradually included into the model at positions corresponding to peaks in the $|F_o| - |F_c|$ electron density map with heights greater than 3σ and at H-bond distance from appropriate atoms. Data collection and refinement statistics are shown in table 3.11. The program PROCHECK (Laskowski *et al.*, 1993) was used to assess the quality of the final structure. Analysis of the Ramachandran plot showed all the residues for

the wild type α3GT complex with UDP and pNPGal to be within the allowed regions of the plot (Figure 3.33).

Table 3.11: Data collection and refinement statistics for WT α3GT-p-Nitrophenyl-Galactoside complex structure

Substrates observed in the crystal structure	Manganese
	UDP
	p-Nitrophenyl-Galactoside
Disordered regions	none
Space group	$P2_1$
Mols/A.S.U	2
Crystal parameters	
a (Å)	45.15
b (Å)	94.22
c (Å)	94.47
α (°)	90.0
β (°)	99.0
γ (°)	90.0
Resolution (Å)	50.0-2.12
R_{symm} (%)[a] (outermost shell)[b]	8.6 (22.2)
Completeness (%) (outermost shell)[b]	94.4 (80.7)
***I/σI* (outermost shell)[b]**	12.4 (5.14)
Total reflections	258899
Unique reflections	44578
R_{cryst}[c]/R_{free}[d] (%)	16.8/20.5
Ramachandran Plot	
% core/allowed	91.3 /8.7
RMSD from ideal	
Bond angles (°)	1.481
Bond lengths (Å)	0.015
Number of water molecules	552
***B*-factor statistics (Å^2)**	
Protein[e]	18.2 (A), 18.0 (B)
Ligand in the active site[e]	13.0 (UDP-A), 33.1 (pNPGal), 13.7 (UDP-B)
	35.8 (pNPGal)
Metal[e]	14.9 (MN1-A), 15.0 (MN1-B)

a $R_{symm}=\Sigma_h\Sigma_i[|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)]$, where *Ii* is the *i*th measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of *I*(*h*).
b Outermost shell are 2.17-2.12 Å .
c $R_{cryst}= \Sigma_h|F_o - F_c|/\Sigma hF_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection *h*.
d R_{free} is equal to R_{cryst} for a randomly selected 2.6% subset of reflections, not used in refinement.
e Temperature factors for individual molecules quoted

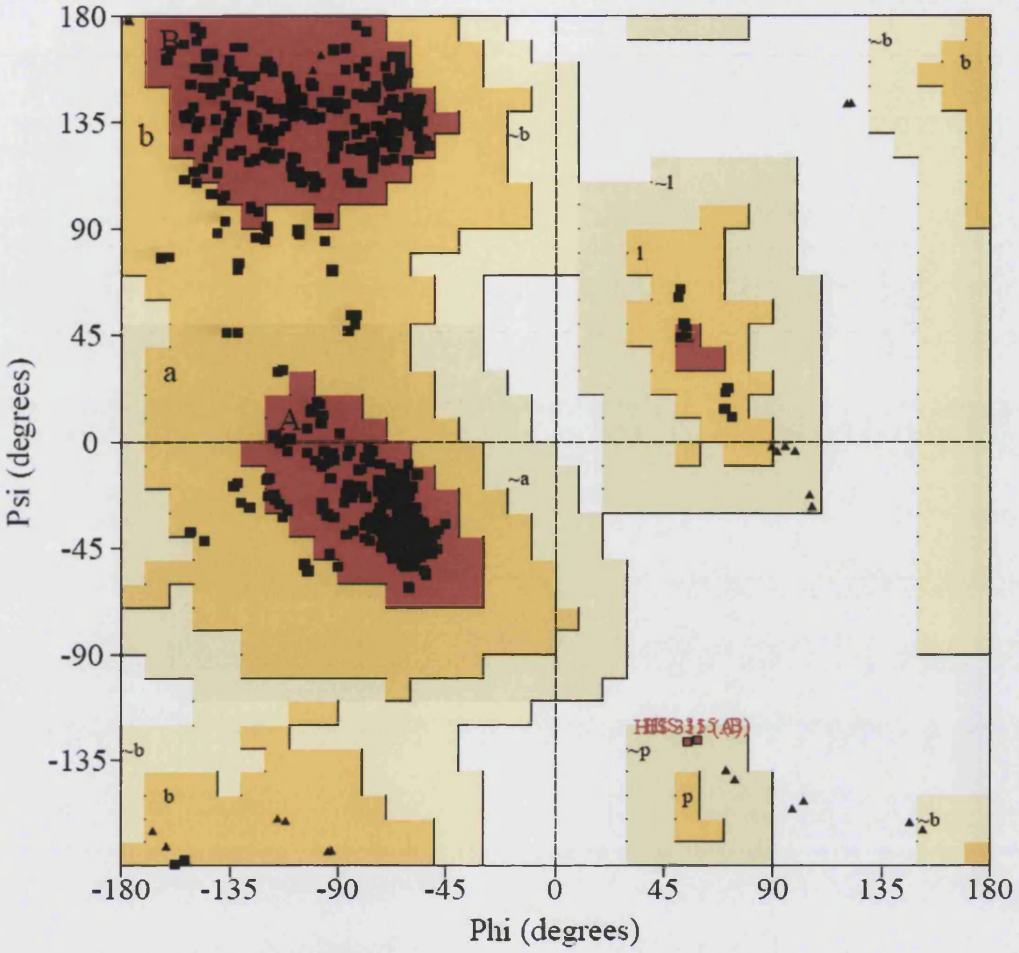


Figure 3.33: Ramachandran plot for the refined structure of wild type α3GT in complex with P-nitrophenyl galactoside (pNPGal) at 2.12 Å resolution. The red coloured regions of the plot (*A, B, L*) represents the residues in the most favoured regions; yellow region (*a,b,l,p*) represents residues in the additional allowed regions; grey/buff coloured regions (*~a,~b,~l,~p*) represents residues in generously allowed regions and white areas of the plot represents the sterically disallowed regions. The glycine residues are represented by triangles; *A/a/~a* represent alpha helices; *B/b/~b* represent beta strands, *L,l,~l* represent left-handed helices; *p/~p* represent epsilon.

Results

p-Nitrophenyl galactoside (pNPGal) binds at the acceptor binding site of α3GT in which the galactose moiety binds similarly to the galactose moiety of the enzyme's natural acceptor substrate N-acetyllactosamine (LacNAc). However binding of the nitrophenyl moiety differs slightly from the binding of the GlcNAc moiety of LacNAc in which the position of the nitrophenyl ring is slightly perturbed compared to the GlcNAc ring of LacNAc (Figure 3.34 and 3.35). The perturbation in the orientation is due to the absence of vicinal hydroxyl groups on the nitrophenyl ring of pNPGal. These vicinal hydroxyl groups on LacNAc make more hydrogen bond interactions with water molecules. In LacNAc, the vicinal hydroxyl, O3 of the GlcNAc moiety interacts directly with NE1 of Trp250. This interaction as well as hydrogen bond interactions with water molecules and the hydrophobic interactions which exist between the rings, facilitate in stabilizing the position of the GlcNAc ring in its parallel stacking orientation to the ring of Trp249 (Figure 3.35 and Figure 3.36). This might explain the higher binding affinity that α3GT has towards its natural acceptor substrate, LacNAc compared to pNPGal. Based on hydrogen bonding calculation of HBPLUS (McDonald and Thornton, 1994), LacNAc has more hydrogen bonding interactions with the surrounding residues compared to pNPGal while pNPGal makes more hydrophobic interactions with the surrounding residues (Figure 3.36). The higher number of hydrogen bonding formation between LacNAc and the binding site residue again is reflected in stronger binding of LacNAc at the active site.

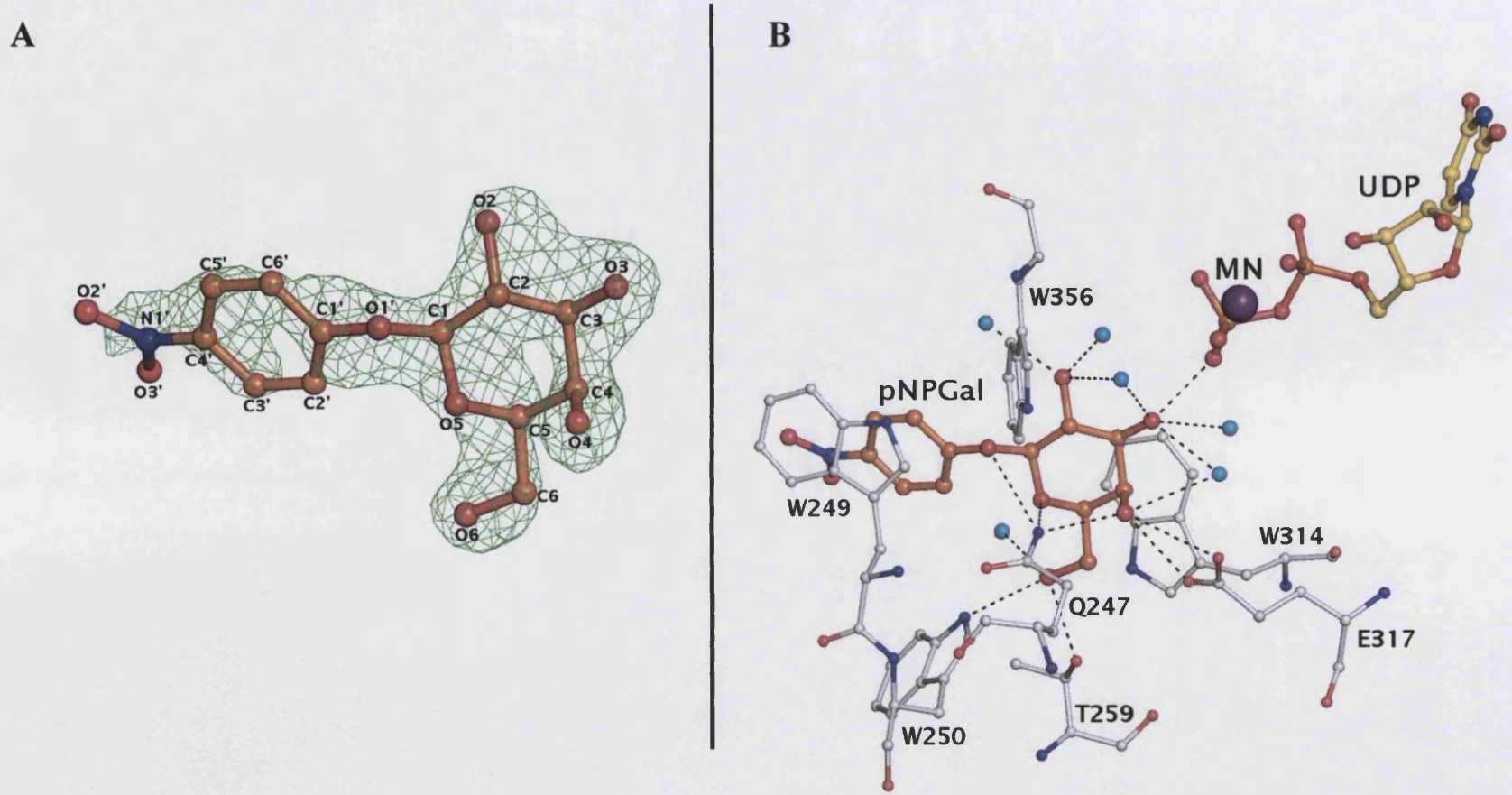


Figure 3.34: Diagram of the Fo-Fc electron density contoured at 3.0 σ level, of pNPGal. Diagram B shows the interactions of α3GT with pNPGal. The protein residues are drawn as *ball-and-stick models*, water molecules appear as *cyan spheres*, and the pNPGal is shown in *orange*. The manganese ion (MN) is shown as a *purple sphere*. H-bonds are indicated by *dashed lines*. The figures were created with pymol.

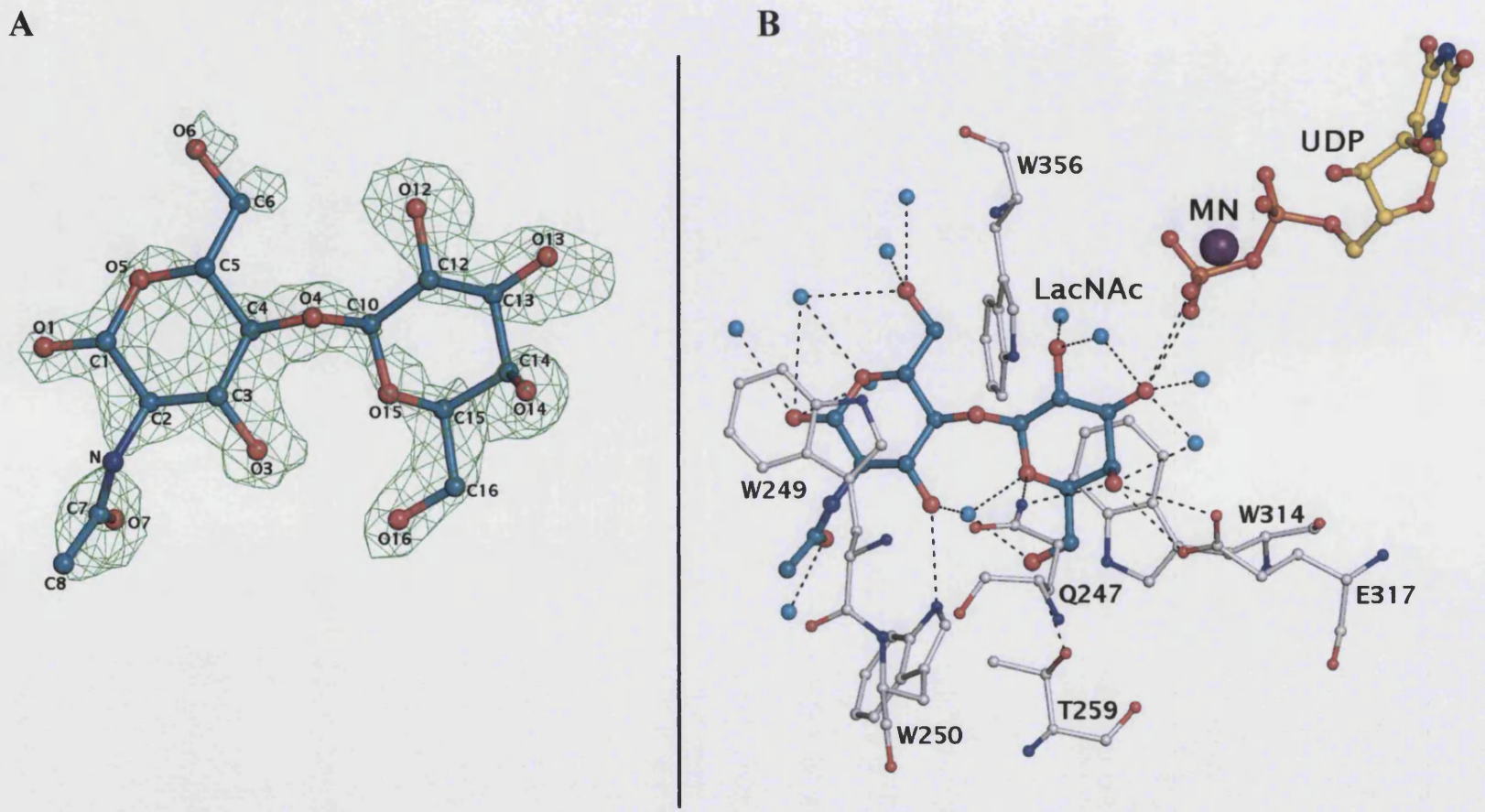
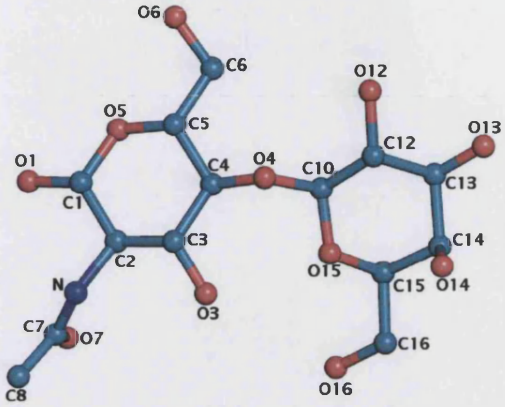
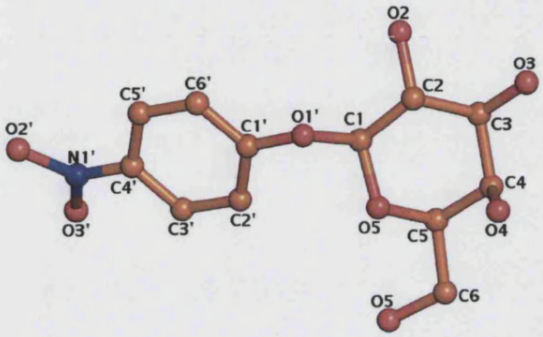


Figure 3.35: Diagram of the *Fo-Fc* electron density contoured at 3.0 σ level, of LacNAc in α3GT. Diagram B shows the interactions of α3GT with LacNAc. The protein residues are drawn as *ball-and-stick models*, water molecules appear as *cyan spheres*, and the LacNAc is shown in *teal*. The manganese ion (MN) is shown as a *purple sphere*. H-bonds are indicated by *dashed lines*. The figures were created with pymol.

Table 3.12: Comparison of the close interactions (between 2.5Å and 3.5Å) between the natural acceptor substrate LacNAc and the alternative acceptor substrate p-Nitrophenyl galactoside (pNPGal) with wild type α3GT

Wild type α3GT•LacNAc complex [PDB code: 1GX4 (7)]			Wild type α3GT•*p*-Nitrophenyl galactoside (pNPGal) complex		
					
Residue or ligand/Atom	LacNAc	Distance (Å)	Residue or ligand/Atom	pNPGal	Distance (Å)
Glu317/OE1	O14	2.83	Glu317/OE1	O4	2.75
Glu317/OE2	O14	3.29	Gln247/NE2	O4	2.85
Gln247/NE2	O14	3.18	Gln247/NE2	O5	2.90
Thr259/OG1	O16	2.85	Gln247/ NE2	O1'	3.39
Trp250/NE1	O3	3.21	Thr259/OG1	O6	2.68
Glu317/OE2	O14	3.29	Glu317/OE2	O4	3.28
Gln247/NE2	O15	2.99	Glu317/OE1	O4	2.73
UDP/O1B	O13	3.31	Trp250/NE1	O6	3.28
UDP/O2B	O13	2.67	UDP/O2B	O3	2.54

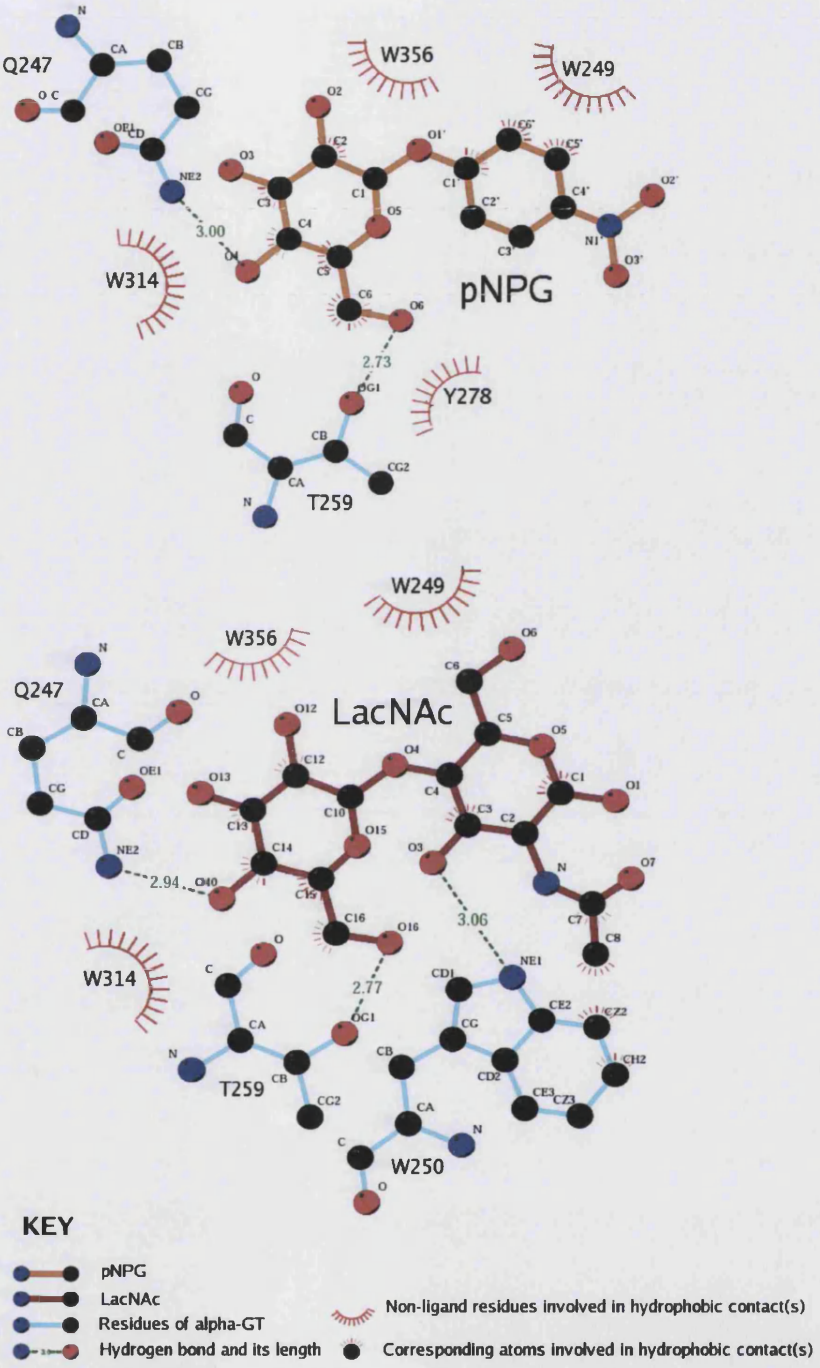


Figure 3.36: Comparison of hydrogen bond interactions and hydrophobic interactions calculated by the program HBPLUS (McDonald and Thornton, 1994).

Discussion

Glycoconjugates which are bound to proteins or lipids play an essential role in many biological processes. Glycoconjugates play an important role in cell signaling and recognition and have been implicated in related processes such as cellular differentiation and immune response. The varied role that glycoconjugates have suggests the power of carbohydrates in therapeutic approaches. For example, the α-1 acid glycoprotein and the IgG molecule both display variations in glycan structure in disease when compared to those on healthy references cells during acute phase response and rheumatoid arthritis respectively (Dwek, 1996; Mackiewicz *et al.*, 1991; Varki, 1993). The tetrasaccharide sialyl Lewis X and related structures are key determinants in the recruitment of lymphocytes during inflammation (Springer, 1990) and furthermore, it has long been known that glycoconjugates on protein carriers are able to induce an antibody response that might protect an organism from infection. α3GT in particular has the potential to be used in cancer therapeutics as the product of the enzyme the α-Gal epitope has been shown to elicit strong immune response in humans. Theoretically, a rationally designed conjugate with a specific ligand at one end that specifically binds cancer cells and an α-Gal epitope at the other end will coat the cancer cells with α-Gal epitopes, and induce anti-Gal mediated immune response. For this reason there is an increased interest in the use of α-Gal epitopes for the treatment of prostate or breast cancer cells. Hence, the study on the enzymatic synthesis of the α-Gal epitope by the enzyme α3GT is an important tool in glycodiversification. Understanding the catalytic mechanism of α3GT by means of structural studies can assist in structure based protein engineering to generate α3GT with new substrate specificity hence able to synthesise novel glycoconjugates for therapeutic use.

Some glycosyltransferases such as α-1-4-galactosyltransferase from *Neisseria meningitidis* LgtC, which transfer galactose from a uridine 5'-diphosphogalactose (UDP-Gal) donor substrate to the axial 4'-OH of terminal lactose containing acceptor substrate has been shown to display acceptor substrate promiscuity. Aside from lactose, LgtC also has also been shown to transfer to a D-galactose, albeit with a markedly reduced efficiency and also to a range of D-aldose

sugars and *myo*-inositol of which some of these sugars possess equatorial hydroxyl groups (Lairson *et al.*, 2006). Further studies showed that LgtC can use *p*-nitrophenyl β-D-glucopyranoside (pNPGlc), which has an equatorial hydroxyl as an acceptor to form a Gal-α-(1-4)-Glc linkage, indicating that pNPGlc adopts the same basic binding mode as lactose (Lairson *et al.*, 2006). This result indicates the possibility that some wild type glycosyltransferases that display a relatively broad specificity can be used as catalysts in a range of useful galactosylation reactions. The use of pNPGlc as an acceptor produced a single regioisomer further suggesting that a strategy whereby appending a conveniently installed and removed substituent onto the acceptor might control its active site binding orientation in a useful manner.

Further studies by Lairson and associates showed that α3GT can utilize both D-galactose and pNPGal as acceptor substrates. The use of D-galactose as an acceptor produced a mixture of regioisomers of disaccharide products whereas the use of pNPGal resulted in exclusive formation of a pure α-(1-3)-linked product (Lairson *et al.*, 2006). This result showed that wild type α3GT has some degree of acceptor substrate promiscuity and has the potential to be used in the assembly of different glycoforms by altering the attachment of activated sugar substrates to variable acceptor substrates. The crystal structure of wild type α3GT in complex with UDP and pNPGal showed that the mode of binding for pNPGal is similar to that of the natural acceptor substrate, LacNAc. α3GT possesses an aromatic side chain (Trp249) that serves as a platform for the binding of the GlcNAc moiety of the LacNAc acceptor. In fact, three dimensional structures of sugar binding sites often show aromatic residues that function to provide a hydrophobic platform onto which a sugar ring face can stack. As with α3GT, LgtC also possesses an aromatic side chain Phe132, positioned to function in such a role, providing a platform on which the sugar ring of the acceptor may bind. The galactose moiety of pNPGal binds in a similar fashion as the galactose moiety of LacNAc. The nitrophenyl group binds differently from the GlcNAc moiety of LacNAc however it still mimicks the stacking interaction of GlcNAc with residue Trp249. Conventional attempts to broaden the substrate specificity of enzymes often involve the modification of both the substrates and the enzyme (Carter and Wells, 1987). Studies carried out by

Lairson and associates showed that by the simple expedient of transiently attaching alkyl or aryl substituent to the acceptor sugar they were able to broaden the substrate specificity of wild type LgtC to allow exclusive formation of single regioisomer products at synthetically useful rates (Lairson *et al.*, 2006). The results presented by Lairson and associates along with the structure of wild type α3GT in complex with UDP and pNPGal presented here highlighted that the acceptor substrate specificity of glycosyltransferases is not quite as rigid as previously believed and this observation has direct implications for their synthetic utility. These results also showed that this substrate promiscuity can be harnessed by modification of the acceptor sugar with a readily removable aromatic substituent that can facilitate a productive and predictable binding mode providing both control of regioselectivity and an increase in catalytic rates (Lairson *et al.*, 2006). The possibility of substrate engineering of glycosyltransferases can provide access to novel glycosidic linkages which may be used for therapeutic purposes.

3.4.4.2 Modelling of the inhibitor U66 with α3GT

During the initial part of my research, one of the main aims was to carry out docking studies of the ligand U66 which was shown to have a high inhibitory activity towards α3GT. Our collaborator, Dr Keith Brew from Florida Atlantic University, Boca Raton (USA) has screened 86 different potential inhibitors which were designed based on the structure of α3GT's natural donor substrate, uridine diphosphate (UDP) (Table 3.13). From the screening studies, they have found that the ligand designated as U66 has the most potent inhibitory activity, with an inhibitory activity within a range of 24 ± 40 μM (Figure 3.37). Docking studies were carried out with U66 and the crystal structure of α3GT (PDB code: 1K4V).

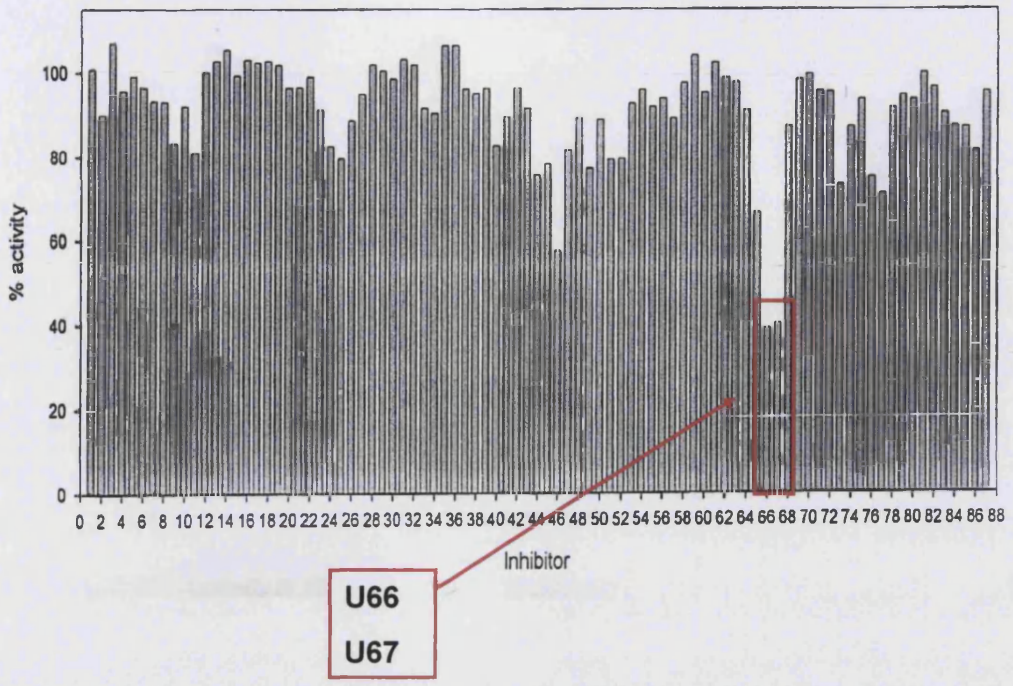


Figure 3.37: Diagram showing the 86 different inhibitors which were screened for inhibitory activity towards α3GT. Inhibitors U66 and U67 (represented in red box) were found to have the most potent inhibitory activity towards α3GT (Brew, K., unpublished results).

Methods

As kinetic studies showed that U66 is close to being a competitive inhibitor with respect to UDP-Gal and LacNAc (data not shown), U66 were built manually in the UDP binding site and LacNAc binding sites. The anthraquinone like residues (Three rings) are predicted to form stacking interactions with tryptophans in the tryptophan pocket and protrude out into solution. The molecular structure of U66 was built using the program 'O'(Jones *et al.*, 1991). The residues used to build U66 were obtained from the HIC-Up server (Kleywegt and Jones, 1998)(http://alpha2.bmc.uu.se/hicup/) and the Cambridge Crystallographic Data Centre (CCDC) (http://cds.dl.ac.uk/cds). Residues that make up the structure of the U66 are as follows (Figure 3.37).

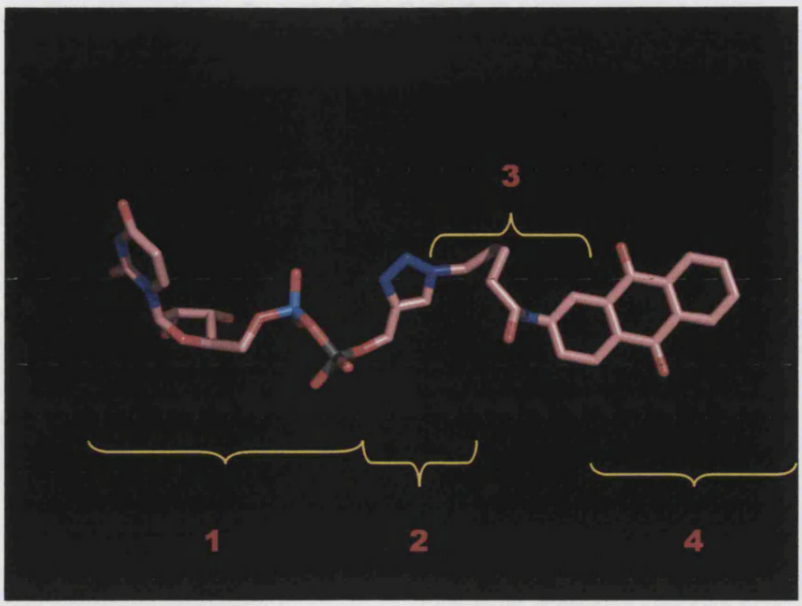


Figure 3.37: structure of U66

1. UDP- derived from the crystal structure of α3GT (PDB code: 1K4V)(Boix *et al.*, 2001).

The rest of the residues were built around the UDP from the α3GT crystal structure (PDB code: 1K4V). Some editing was carried out in order to obtain the correct atoms for the inhibitor:

2. Triazole ring (Mangani and Liljas, 1993): HIC-Up server
3. Pentanal (Nevskaya *et al.*, 2000): HIC-Up server
4. Anthraquinone (Henry and Lonsdale, 1952): Crystal web (CCDC)

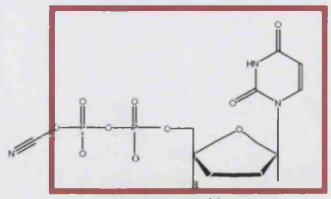
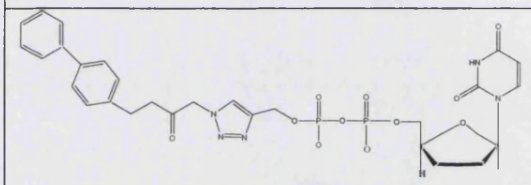
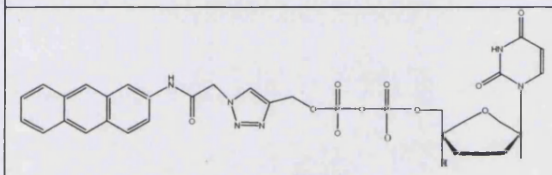
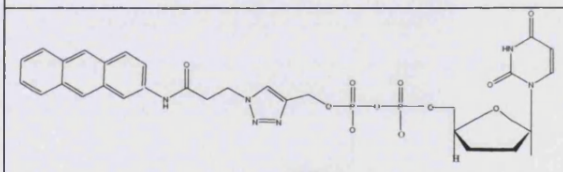
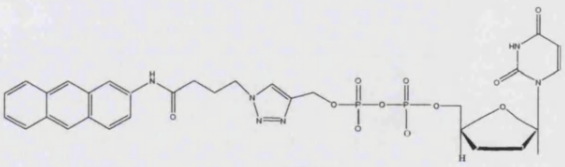
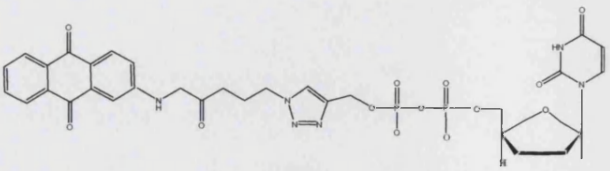
Inhibitors	K_i (μM)
 UDP-Alkyne	2,498 ± 790
 U21	110 ± 14
 U43	51 ± 14
 U44	57 ± 14
 U45	47 ± 8
 U66	24 ± 40

Table 3.13: Some of the ligands that have been screened for inhibitory activity towards α3GT. The common structure for all 86 inhibitors is shown in the red box.

Step 1: Manually build inhibitor

The modelled complexes were refined using the program CNS (Brunger *et al.*, 1998) and the topology and parameter files for refinement were obtained from the HIC-Up server (Kleywegt and Jones, 1998).

Step 2: Docking studies to verify manually built model

Autodock (Goodsell and Olson, 1990):

Docking studies were carried out using the program AUTODOCK 3.0. Autodock carries out Grid-based empirical flexible docking via Monte Carlo search and incremental construction (www.scripps.edu/pub/olson-web/download.html). Autodock uses a kinematic model for the ligand similar to the one illustrated in figure below. The ligand begins the search process randomly outside the binding site and by exploring the values for translations, rotations and its internal degrees of freedom, it will eventually reach the bound conformation. Distinction between good and bad docked conformations was carried out by the scoring function (Teodoro *et al.*, 2003).

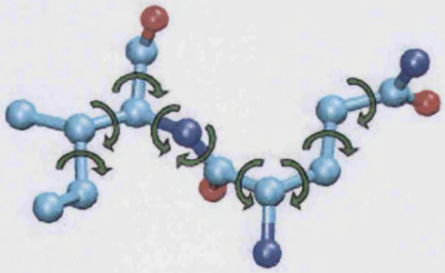


Figure 3.38: A drug molecule. Spheres represent atoms and bonds connecting them are represented by sticks. Curved arrows represent the rotatable degrees of freedom around bonds (Teodoro *et al.*, 2003).

Partial atomic charges were assigned to α3GT and U66 using the program SYBYL 6.8 (Ryckaert *et al.*, 1997). For α3GT, all water molecules were removed and the atomic solvation parameters were assigned to the α3GT atoms using the ADDSOL utility. The grid maps around the protein binding site were calculated using AUTOGRID 3.0 with 80 x 80 x 80 points, and a grid spacing of the default

0.375 Å (roughly a quarter of the length of a carbon-carbon single bond). The grid centre used was initially 6.122, 13.016, -4.651 based on the manual-docking model. The grid centre is usually set by default based on the coordinates for the center of the ligand. Torsion angles that were to be explored during the docking were defined using 'deftors' (a program which assigns root atoms and torsions to a molecule). The UDP moiety of the inhibitor was set as the root or fixed reference point from which the rest of the atoms will flex. This is because it is believed that if the inhibitor was to bind at the predicted binding site, the UDP moiety would fit into the binding site in the same conformation as it is in the crystal structure. Torsions angles were mainly centered around the short hydrocarbon chain connecting the triazole ring with the anthraquinone-like residue. The Autodock run were not successful and did not place the inhibitor in the manually docked site. Changing the docking parameters such as the number of runs, the grid center, root residues and torsion angles were still not successful in placing U66 at the expected binding site. At this point it was decided to move to another docking program namely Hex.

Hex (Ritchie and Kemp, 2000)

Hex (http://www.csd.abdn.ac.uk/hex/) is an interactive molecular graphics program for calculating and displaying feasible docking modes of pairs of protein and DNA molecules. Hex can also calculate small-ligand/protein docking (provided the ligand is rigid), and it can superpose pairs of molecules using only knowledge of their 3D shapes. The main thing, which distinguishes Hex from other macromolecular docking programs and molecular graphics packages is its use of spherical polar Fourier correlations to accelerate the docking and superposition calculations. An important difference between Hex and Autodock is that Hex is much more user friendly as it uses a graphics interface. Another difference is that in Autodock, the docking determination takes into account the flexibility of the ligand through the calculation of torsion angles within the ligand molecule whilst Hex is based on the principle that the ligand is rigid. Initial runs to dock the inhibitors to α3GT using the Hex default parameters were not successful. Consequently, U66 was manoeuvred in

close proximity to the predicted binding site (i.e: close to the tryptophan pocket). The Euler angles for the orientation were as follows: α- 97, β- 36, γ- 63° and an intermolecular separation of 26.4 Å. Only chain B of the α3GT molecule was used. Prior to docking, as with Autodock, the water molecules in α3GT were all removed. The default mode, which is full rotation, was used as the docking search mode. In this mode, the receptor is held fixed and the ligand is translated about the receptor using the Ligand Range and Samples angular parameters and the Distance Range and Step parameters to generate different ligand poses about the receptor. The resolution was specified at the default settings of an initial Steric Scan at N=16, followed by a Final Search at N=25, using just the steric contribution to the docking energy. In this mode, about all but the top 20,000 orientations are discarded after the Steric Scan. The docking was performed using the angular search of 180° angular cones centred on the intermolecular axis. Using all these parameters, the inhibitor was placed into the predicted binding site and gave the first docking solution with the lowest free energy value of -390.48 (Figure 3.39). Subsequent docking trials showed that the U66 would be docked in the predicted binding site if the inhibitors were oriented in close proximity to the tryptophan pocket prior to docking.

Although the Hex solutions are in agreement with the manually docked model, the model should be viewed with some trepidation as the results were only obtained when the inhibitor was placed in very close proximity to the predicted binding site. This could reflect a bias in the search towards a pre selected binding site, therefore compromising the validity of the results.

Results

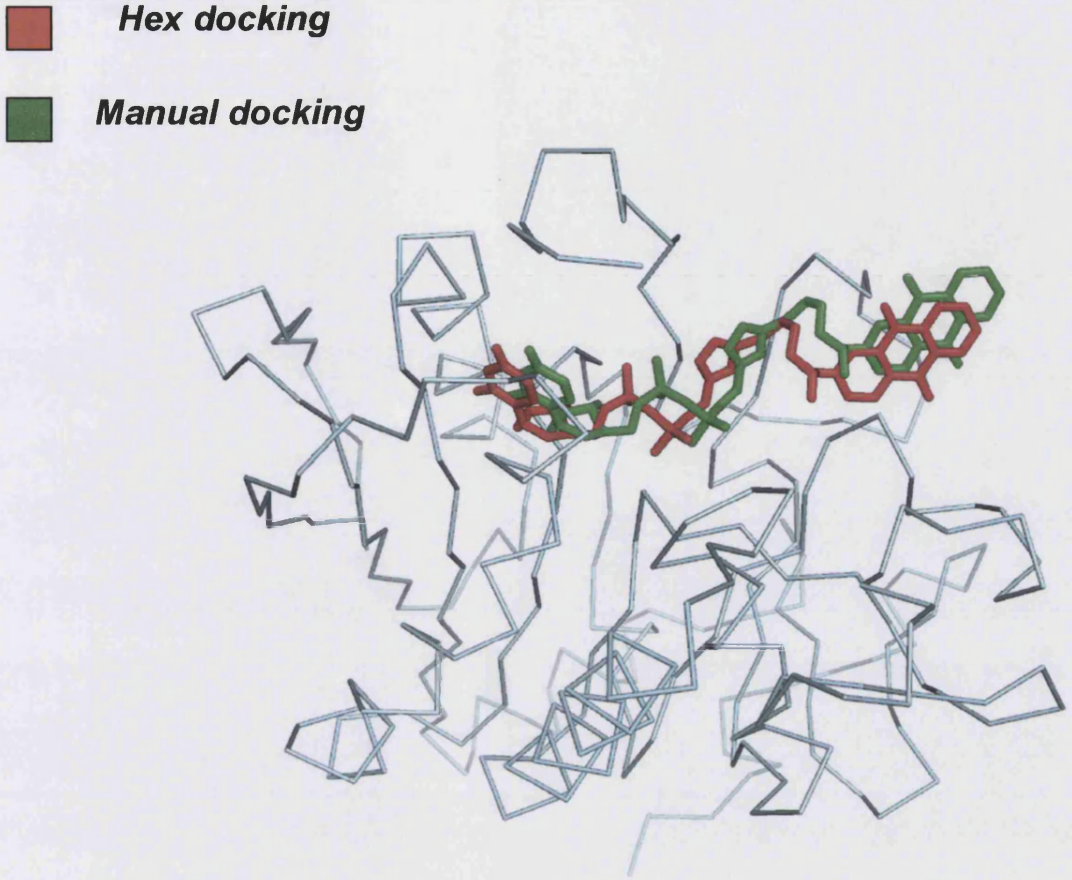


Figure 3.39: Diagram showing the manually docked U66 and the result obtained via docking using the Hex program. The manual docking position is represented in red while Hex docking is in green. The Hex docking result is in good agreement with the manually docked position and gave a low free energy value of -390.48.

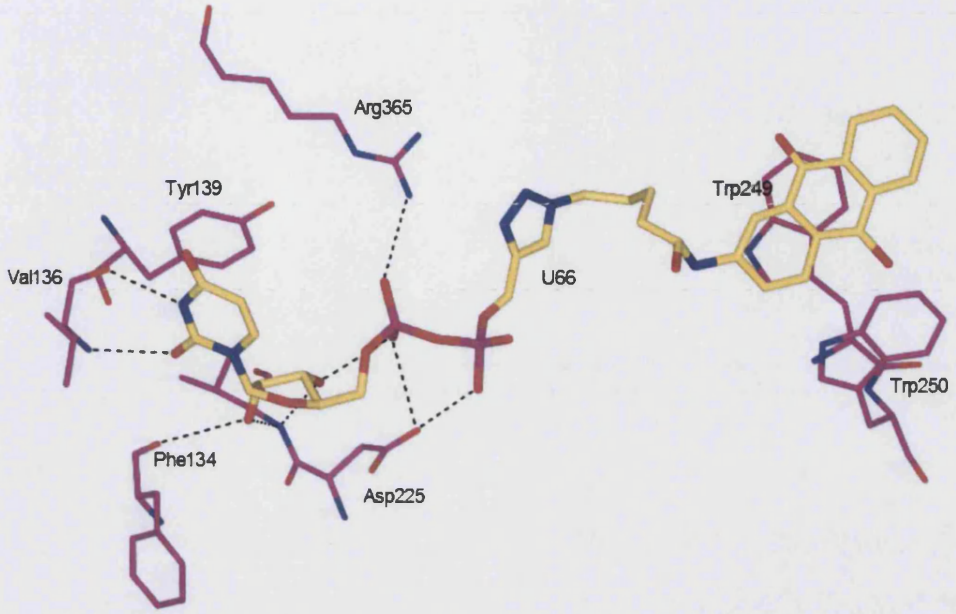
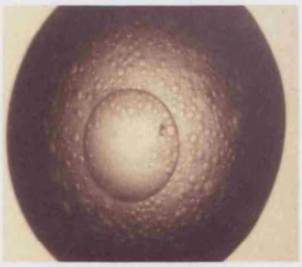


Figure 3.40: Potential inter-atomic interactions between the docked ligand, U66 (represented in yellow) with the active site residues of α3GT (represented in magenta). The black dashed line represents hydrogen bonds. Stacking interactions between the anthraquinone-like tail of U66 with the tryptophan rings (residues 249 and 250) are predicted to stabilise the conformation of U66.

U66	α3GT	Distance (Å)
O1A	Asp225-OD2	3.22
O1B	Asp225-OD2	2.76
O2*	Val226-N	3.4
O3*	Val226-N	3.5
O2*	Phe134-O	3.3
O2	Val136-N	3.16
N3	Val136-O	2.9
N2A	Tyr361-OH	2.45
N5A	Tyr361-OH	3.36

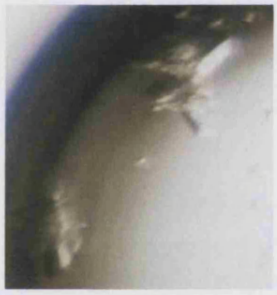
Attempts to obtain crystal structure of α3GT in complex with U66

One of the main objectives of this research initially was to co-crystallise wild type α3GT with the inhibitor U66, however, initial attempts to co-crystallize wild type α3GT with the inhibitor U66 were unsuccessful (data not shown). On addition of U66, α3GT precipitated and it was concluded that the best way to overcome this problem was to grow α3GT crystals without substrates or co-factor (apo) and then carry out soaking experiments with U66. Initial soaking experiments of wild type α3GT –UDP co-crystals with U66 had shown that U66 was not able to displace UDP from the active site (data not shown). At this point we abandoned further efforts to co-crystallise U66 with α3GT and sought to obtain apo crystals of the enzyme for use in soaking experiments with U66. We attempted to obtain crystallisation conditions for the apo forms of the wild type enzyme and the Arg365Lys mutant. Apo wild type α3GT was successfully crystallised in the tetragonal form previously reported by Zhang and co-workers (Zhang *et al.*, 2003) Crystallisation of α3GT mutants was carried out using the vapour diffusion method. Initially, crystallisation screens of the Arg365Lys mutant in its 'apo' form did not produce any viable results. The crystallisations results for the screens were mostly bleak with the results mainly being clear drops and heavy precipitation, with the latter predominating. It was observed that condition number 12 from Structure screen I consistently produced phase separation (Figure 3.41) for the Arg365Lys mutant. Phase separation or also known as protein droplets can be a positive indicator of a potential crystallisation condition and it is sometimes likely that crystals would grow from phase separation (Bergfors, 1999). Based on this knowledge, streak seeding was carried out. The same condition was repeated and the new drops containing the Arg365Lys mutant were streak seeded from the phase separation. Crystalloids appeared after a week of incubation and streak seeding was carried out subsequently until the crystal size and morphology improved (Figure 3.41). The condition for crystallisation was also refined and the most optimum condition was established to be in the presence of 2M sodium acetate and 0.3M sodium cacodylate pH 6.5.



Phase separation. A similar phenomena occurred when Arg365Lys variant (apo) was crystallised in Structure Screen I condition number 12: 1.4M sodium acetate, 0.1M sodium cacodylate, pH 6.5

Crystalloid of Arg365Lys variant formed a week after streak seeding from phase separation. The crystalloids grew in the form of dense clusters and are not useful for data collection



After subsequent streak seeding from crystalloids and refinement of crystallisation condition, Arg365Lys variant crystal morphology and size improved significantly. Crystallisation condition: 2M sodium acetate, 0.3M sodium cacodylate, pH 6.5

Figure 3.41: Diagram showing gradual improvement of mutant Arg365Lys 'apo' crystal morphology via streak seeding. This crystal diffracted to 3.3 Å and crystallised in orthorhombic space group $P2_12_12_1$.

The same condition was used to co-crystallise Arg365Lys mutant with UDP-Gal and manganese. These attempts produced small crystals of bipyramidal morphology (Figure 3.42).

Figure 3.42: Crystal of Arg365Lys variant grown in the presence of UDP-Gal and manganese Crystallisation condition: 2M sodum acetate and 0.3M Na sodium cacodylate, pH 6.5. The crystal diffracted to 3.7 Å in tetragonal space group (Typical crystal size: 0.06 X 0.07 X 0.18 mm).

These apo crystals however did not diffract well hence are not ideal crystals to be used for soaking experiments. Initial attempts to soak these crystals in U66 were unsuccessful and due to time constraints further attempts at obtaining a crystal structure of α3GT in complex with U66 were stopped so that the research could move on to other areas of study

Discussion:

Selective glycosyltransferase inhibitors would be valuable tools for probing the functions of glycosyltransferases and their products. With few exceptions, however, selective glycosyltransferase inhibitors do not exist. The design of inhibitors for α3GT is important in the relation to the problems of xenotransplant rejection caused by the recipient's anti-Gal antibody reaction to the donor's α-Gal epitope. Several inhibitors of α3GT have been reported and some of them is an 1-*N*-iminosugar based UDP-galactose analogue which was found to be a potent and selective inhibitor of α3GT but not of the β-1,4 galactosyltransferase. Azasugars have also been designed to inhibit the enzyme activity by forming a strongly associated complex in the active site via favourable electrostatic interaction (hydrogen-bonded ion pair) in a manner similar to inhibitors of retaining glycosidases (Compain and Martin, 2001). Studies by Takayama and associates showed that 1-deoxygalactonojirimycin which is a potent inhibitor of ceramide glucosyltransferase showed no inhibition towards α3GT

nor towards β-1,4 galactosyltransferase (Takayama *et al.*, 1999). These results indicate clearly that the position of the imino-group is crucial for the recognition of the piperidinols by the enzyme. The design of inhibitors of α3GT is still at an early stage but even though we did not manage to obtain a crystal structure of the enzyme in complex with U66, the modeling and docking studies seem to suggest that U66 binding spans both the donor substrate and the acceptor substrate site. The inhibitory nature of the compound may be attributed to this mode of binding as it obstructs both substrate binding sites hence preventing the natural substrates from binding at the active site. Designing a selective inhibitor of α3GT will require greater understanding of the catalytic activity of the enzyme as well as attempts to obtain crystal structures of α3GT in complex with inhibitors in order to have a detailed understanding of their binding mode. The information acquired would make it possible to prepare a focused library of compounds to identify better and more potent inhibitors.

CHAPTER 4

Crystallisation of Canine Forssman Synthase (FS)

4.1 Introduction

Glycolipids are carbohydrate-attached lipids. Their role is to provide energy and also serve as markers for cellular recognition. They occur where a carbohydrate chain is associated with phospholipids on the cell membrane surface. The carbohydrates are found on the outer surface of all eukaryotic cell membranes. They extend from the phospholipid bilayer into the aqueous environment outside the cell where they act as a recognition site for specific chemicals as well as helping to maintain the stability of the membrane and attaching cells to one another to form tissues.

Microbial pathogens have long co-evolved with their eukaryotic hosts and this has resulted in a remarkable specificity in their interactions. As a consequence of this, pathogenic organisms have a restricted ability to cause disease to a limited number of species. The early stage in the development of pathogenesis of most infectious diseases is microbial adherence to the host cells (Xu *et al.*, 1999). As a result, the capacity for adherence has become a major determinant of the host range available to a given pathogen. Due to their role in cell recognition, glycolipids are usually the initial attachment site for several pathogenic agents such as bacteria, viruses and bacterial toxins. These pathogens are known to recognize host glycolipids as their initial attachment site (Hansson *et al.*, 1985; Hansson *et al.*, 1984; Karlsson and Stromberg, 1987; Svanborg Eden *et al.*, 1983). Hence, the susceptibility to infectious diseases is impacted greatly by factors regulating glycolipid synthesis.

Glycolipids are composed of a variable carbohydrate moiety attached to a ceramide backbone and can be found in almost all eukaryotic cells. Hundreds of distinct glycolipids have thus far been described (Basu *et al.*, 1987). The variability in glycolipid structures takes place primarily in the carbohydrate moiety due to differences in the number, type, or anomeric linkage between sugar residues (Stults *et al.*, 1989; Xu *et al.*, 1999). However, each cell type will only express glycolipids which contain a limited combination of all possible carbohydrate structures. This cell lineage-specific pattern of glycolipid expression is tightly regulated during

cellular differentiation and development and also varies between species (Hakomori, 1986; Monner and Muhlradt, 1993; Sadahira *et al.*, 1991; Sandhoff and van Echten, 1993).

The enzyme globoside α-N-acetylgalactosaminyltransferase (EC-Number 2.4.1.88) or also known as Forssman synthetase (FS) is responsible for the synthesis of Forssman glycolipid (FG; Forssman antigen) which is a member of the globoseries glycolipid family, all of which have in common a core galactosyl-(α1,3)galactose moiety (Xu *et al.*, 1999). FS catalyzes the transfer of *N*-acetylgalactosamine from its nucleotide donor, UDP-*N*-acetyl-D-galactosamine to the acceptor, *N*-acetyl-D-galactosaminyl-1,3-D-galactosyl-1,4-D-galactosyl-1,4-D-glucosylceramide to form its product the Forssman glycolipid (FG), *N*-acetyl-D-galactosaminyl-α-1,3-N-acetyl-D-galactosaminyl-α-1,3-D-galactosyl-α-1,4-D-galactosyl-α-1,4-D-glucosylceramide. An interesting feature of FG is that unlike many other mammalian species, human cells do not normally produce it but instead produce the precursor glycolipids globotriosylceramide and globoside (GbO4). In humans and some species, these precursors serve as attachment site for pathogenic agents. In other species, it is likely that the modification of these glycolipids, such as the addition of *N*-acetylgalactosamine to create FG, alters the adherence of pathogenic organisms hence modifying host susceptibility to infectious diseases. A study by Elliot and co-workers (2003) described a molecular and genetic relationship between FG expression and Shiga toxin (Stx) susceptibility. In this study, they isolated the FS cDNA from human, canine, and murine cells and found that whereas the murine and canine FS genes express a functional enzyme, the human FS cDNA was found to express a protein that lacks FS activity, despite a high degree of sequence identity with the enzymatically active murine and canine FS genes. In order to examine the relationship between FG expression and Stx susceptibility, Vero cells were transfected with the three FS orthologues or a vector control. Complementation with the human FS cDNA had no effect on Stx susceptibility, whereas stable expression of the canine and murine FS resulted in markedly decreased susceptibility to toxin. Among individual cells, an inverse correlation between FG expression and Stx binding was demonstrated. Moreover,

only strongly FG-reactive cells were capable of growing in the presence of Stx. These cells were found to have high levels of FG expression and a correspondingly diminished GbO3 content. Based on these results, Elliot and co-workers concluded that expression of a functionally active FS modifies Stx receptor glycolipids to FG and results in markedly decreased susceptibility to toxin. They further speculated that inactivation of the FS gene during primate evolution may account for the marked susceptibility of human cells to Stx (Elliott *et al.*, 2003). Studies by Xu and co-workers showed that the inability to produce FG results from mutations within the putative FS catalytic domain. However, widespread tissue expression of the FS mRNA seems to suggest that although the human FS is not active it may still retain some biological function and the inactivity of the enzyme in humans may be due to 2 possible reasons. The first reason might be due to FS acquiring a novel enzymatic activity or it may be because the remaining biological functions in humans do not require glycosyltransferases activity. Xu and co-workers moved on to compare the human orthologue of FS with that of α3GT. α3GT is similar to FS in that the α3GT gene is expressed in many mammalian species however it lost the ability to encode glycosyltransferase activity only recently during primate evolution. Interestingly, whereas the human FS retains 83% amino acid identity with the canine orthologue and is expressed widely in human tissues, the α3GT is not expressed in human cells, and the processed pseudogene contains multiple frameshift mutations (Galili *et al.*, 1988; Xu *et al.*, 1999).

The study reported in this thesis concentrated on the structural studies of these 2 family 6 glycosyltransferase with the bulk of the study concentrated on structural studies of several α3GT mutants in order to further understand the catalytic activity and specificity of this enzyme. The important role that FG plays in host-pathogen interactions and by implication, the important role that FS has to play in synthesizing FG makes it crucial for us to understand the structural basis for the catalytic activity of FS. It is in this context that we also sought to elucidate the crystal structure of FS in the hope of understanding structurally its catalytic mechanism.

Methods

Forssman synthase (FS) enzyme was cloned, expressed and purified by Percy Tumbale in our collaborator Professor Keith Brew's laboratory, Florida Atlantic University, USA. The pure enzyme was supplied to us in buffer containing 20mM Mes-NaOH, pH 6.5 and 50% Glycerol at a concentration of 0.48 mg/ml. The protein needed to be buffer exchanged into a buffer containing 20mM Mes-NaOH, pH 6.5 and 10% glycerol and then concentrated to 5-8mg/ml (Abs_{562} between 0.12-0.22) prior to crystallisation set ups. The method used for this step is as described previously in Chapter 3. During the buffer exchange and concentration process, quite a large amount of protein was lost and due to the small amount of protein available, this has put a limitation on the number of crystallisation screens that could be carried out. Described here are the optimisation attempts which had been carried out in light of the limited resources that were available.

Crystallisation was carried out using the hanging drop vapour diffusion method by mixing 1µl protein in the presence of 10mM UDP-GalNAc and 10mM $MnCl_2$ with 1µl of well solution and suspended above 0.5-0.8 ml reservoir. Crystallisation screens were also carried out in the absence of UDP-GalNAc. The drops were incubated at 4°C, 16°C or 22°C and the well solution to protein ratio was also varied in hope of obtaining diffraction quality crystals. Crystallisation trials were initially carried out using published conditions for the human ABO blood group enzymes however, this condition resulted in protein precipitation hence, the focus moved on to commercial crystallisation screens such as the Hampton PEG/ION screen, Hampton Malonate Screen, Hampton Ammonium sulphate screen, Molecular Dimensions Structure Screen I and II, Hampton Crystal Screen Lite, Memstart Kit Screen and Memsys Screen.

4.3 Results

Initial screening was carried out with the commercially available screens at the maximum concentration of protein (8 mg/ml) and incubated at 16°C. These screens represent a hundred conditions which have commonly given crystals in previous experiments. These commercial screens produced many drops with heavy amorphous precipitate suggesting that either the protein or the precipitant concentration was too high in most cases. The screens were carried out a second time with a lower protein concentration of 5 mg/ml and subsequent screens were carried out using this protein concentration. Initial screens at this concentration using the PEG/Ion Screen showed precipitation in all the conditions hence a second crystallisation set up was carried out by halving the precipitant concentration of all the conditions from the PEG/Ion screen. This resulted in most drops having a small amount of precipitates, some of which were more crystalline in nature. Several other screens such as the Molecular Dimensions Structure Screens and Hampton Crystal Screen Lite gave promising results with several conditions producing crystalloids and microcrystals (Table 4.0). However the morphology of these crystalloids and microcrystals were the same in which they produce very fine needle crystals growing in clumps (Figure 4.0). Several optimisation screens based on these initial hits were carried out along with streak, macro and micro seeding in the hope of improving the crystal morphology. The optimisation screens involved varying the protein concentration, precipitant, pH range and also the addition of additives as shown in Table 4.0. Variations on these conditions did not produce crystals with improved crystal morphology. The needle crystals improved tremendously in length (there were several instances in which the needles grew to span the whole length of the drop) however they were still very thin and were still growing in clumps hence were not suitable for data collection. At this point the aim of the optimisation trials was to produce a single needle crystal which could be fished out of the drop easily enough using a cryo loop and used for data collection.

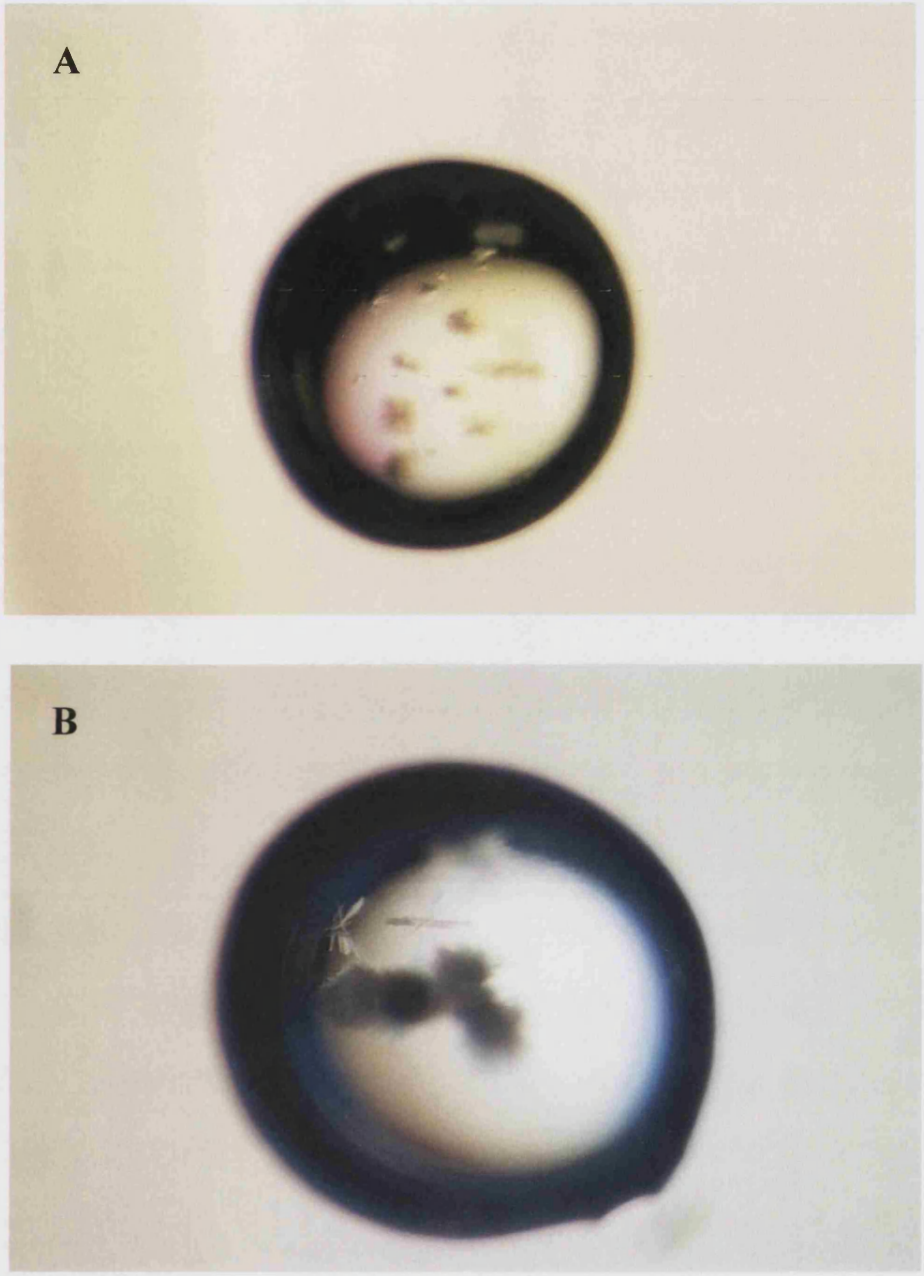


Figure 4.0: A: initial crystalloids obtained from structure screen I/II, PEG/Ion screen and crystal screen lite. B: slight improvement in size of the crystal after optimisation however, the crystals are still very fine needles growing in clumps resembling sea urchins hence, are not suitable for data collection.

Following the lack of success with the optimisation trials to produce diffraction quality crystals, membrane protein screens from Molecular Dimensions were tried: Memsys Kit and Memstart Kit. Interestingly, several conditions (Table 4.0) from these screens produced crystalloids within 2 hours of incubation which is much faster than conditions from the structure screens and Hampton Research screens which normally would take between 2-4 weeks. These crystallisation conditions interestingly reflected the crystallisation condition of human ABO blood group enzymes which had been previously tried for FS crystallisation without much success. Both the human ABO blood group enzyme and the conditions from the Memstart Kit Screen (conditions number 2 and 8) use ADA/NaOH as a buffer and contain the salt Ammonium sulphate albeit at different concentrations. Although the conditions from the Memstart Kit Screen looked promising, the morphology of the crystalloids were still similar to that of previous conditions and the challenge is to improve the conditions in order to obtain a single diffraction quality crystal. As the major difference between the conditions of the human ABO blood group enzyme and the conditions from the Memstart Kit Screen is the pH of the buffer, preliminary optimisation trials were carried out by varying the pH of the ADA buffer in order to find out the pH boundary of FS. This was important as it would give a clearer idea of which pH range to pursue for crystallisation. Crystallisation set ups were carried out using condition number 2 from the Memstart Kit Screen but with two different pH, 6.5 (original pH from Memstart Kit) and 7.5 (pH for the crystallisation condition of human ABO blood group enzyme). The concentration of the precipitant which is ammonium sulphate was also varied between the ranges of 0.8M to 1.5M. Results from this optimisation trial showed that FS is more stable at pH 6.5 and tends to form amorphous precipitates at pH 7.5 which explains why repeating the condition of the human ABO blood group enzyme for FS crystallisation met with failure. The results also showed that an ammonium sulphate concentrations of less than 1.2M showed the best crystal improvement while concentration above 1.2M produced amorphous precipitates. These results provided a clear and defined route for subsequent optimisation trials. As the results from the membrane protein screens were obtained at the time of writing, further optimisation trials were put on hold.

Crystallisation screen	Condition	Optimisation
Molecular Dimensions Structure Screen I	*Number 7:* 30% PEG 4000 0.1M Sodium citrate, pH 5.6 0.2M Ammonium acetate	Fine needles growing in clusters
	Number 12: 1.4M sodium acetate trihydrate 0.1M sodium cacodylate , pH 6.5	Fine needles growing in clusters
	Number 19: 19% PEG 8000 0.1M sodium cacodylate, pH 6.5 0.2M zinc acetate	Fine needles growing in clusters
Molecular Dimensions Structure Screen II	*Number 20:* 1.6M Magnesium sulphate heptahydrate 0.1M MES, pH 6.5	Fine needles growing in clusters
Molecular Dimensions Memstart Kit	*Number 2:* 0.1M ADA/NaOH, pH 6.5 1M Ammonium sulphate	To be optimised further
	Number 8: 0.1M lithium sulphate hydrate 0.1M ADA/NaOH, pH 6.5 1M Magnesium sulphate hydrate	To be optimised further
Molecular Dimensions Memsys Kit	*Number 10:* 0.1M MES, pH 6.5 0.1M sodium chloride 0.1M lithium sulphate 30% PEG 400	To be optimised further
Hampton Research PEG/Ion Screen	*Number 45:* 0.2M lithium citrate tribasic tetrahydrate 20% PEG 3350	Fine needles growing in clusters
	Number 46: 0.2M sodium citrate tribasic dehydrate 20% PEG 3350	Fine needles growing in clusters
Hampton Research Crystal Screen Lite	*Number 39:* 0.1M HEPES sodium, pH 7.5 2% v/v PEG 400 1M Ammonium sulphate	Fine needles growing in clusters
Hampton Research Crystal Screen Lite	*Number 39:* 0.1M HEPES sodium, pH 7.5 2% v/v PEG 400 1M Ammonium sulphate	Fine needles growing in clusters
Hampton Research Malonate Screen	none	N/A
Hampton Research Ammonium Sulphate screen	none	N/A

Table 4.0: Conditions from commercial screens which produced crystalloids or microcrystals and the outcome of optimisation trials of these respective conditions.

Discussion

Crystallisation of proteins requires obtaining the right set of conditions to supersaturate the protein solution in the most favourable way for crystallisation. The challenge in crystallisation is to identify these sets of conditions in order to obtain a diffraction quality crystal as there are numerous factors which can influence the crystallisation ability of a particular protein. Some of the factors that affect the solubility of the protein, and therefore its saturation level are protein concentration, precipitant concentration and type, pH and temperature. These conditions are the easiest to vary in crystallisation experiments hence would usually serve as the variable factors of most optimisation trials. In designing crystallisation trials, it is often hoped that the preliminary commercial screens will produce some form of preliminary crystals that can be optimised by adjusting the aforementioned parameters. From the commercial screens that were tested on FS, there were 10 conditions which showed promising results. However optimisation of 7 of these conditions from the commercial screens of Molecular Dimensions and Hampton Research did not result in any marked improvement of the crystal size but interestingly these conditions showed that FS seemed to crystallise in conditions containing acetate ions and/or citrate ions (Table 15) indicating that these ions can be favourable additives for FS crystallisation.

Subsequent crystallisation trials using the membrane protein screens from Molecular Dimensions resulted in three conditions which produced more promising looking crystalloids and microcrystals. Interestingly enough two of these conditions contain ADA buffer which also contains an acetate component. Preliminary optimisation trials indicated that pH 6.5 and an ammonium sulphate concentration of below 1.2M are most favourable for FS crystallisation. Hence, future optimisation trials may involve fine grid screens of the pH and also concentration of the ammonium sulphate precipitant. Apart from that, seeding techniques should also be applied in the crystallisation experiments especially using the streak seeding and microseeding techniques. Additives (Chapter 3, Table 3) should also be included in the crystallisation trials paying particular attention to the incorporation of acetate

and citrate ions. Another way to explore the crystallisation space more effectively is to set up more screens however, as FS was supplied at low concentration (0.48 mg/ml) and in 50% glycerol this may be difficult. Buffer exchange and concentration of FS prior to setting up crystallisation means that a considerable amount of protein is lost before hand hence this limited amount of protein may hinder larger scale crystallisation trials. If the expression and purification step for producing FS can be improved to produce FS protein at higher concentration, this can minimise the loss of protein as it will shorten the length of the buffer exchange and concentration process. With larger amounts of available protein, more crystallisation trials can be carried out.The use of robots for setting up large trials, and the improvement of computerised crystal identification techniques to observe these trials can also improve the chances of finding suitable conditions. In the absence of a robot, experiments must be designed as well as possible to sample the crystallisation space whilst conserving protein and time required to set up the experiments. The strategies used in these experiments became more effective throughout the trials as experience was gained in designing trials and scoring drops for their precipitate quality. However, due to time constraints there is still a large proportion of the crystallisation space left unexplored. Although these experiments have yet to result in the production of a diffraction quality crystal, they have gone some way towards identifying several leads which can be used for future screens.

Conclusions and Future Work

The reported structures of the C-terminal mutants of α3GT have identified key regions which undergo conformational changes on binding of substrates. From the structures it is evident that the loops encompassing the substrate binding sites are important in catalysis, providing an ideal environment for optimal binding of both the donor and acceptor substrates. These studies also identified a second manganese bound in the active site of the enzyme. This is in correlation with the kinetics studies which indicated that there are two binding sites for the manganese ion (Brew, K *et al.*, unpublished results). The second and weaker binding site for the manganese ion is in close proximity to the Glu317 residue and overlaps with the putative position of the trisaccharride product. As the binding of this second manganese occurs only after the release of the product, it is postulated that this second manganese functions to displace the UDP group after trisaccharide product release in order to allow a new UDP-Gal to bind at the donor site hence maintaining a continuation of the catalytic cycle. It is expected that the second manganese facilitates the release of UDP by disrupting the interactions of the C-terminal residues with UDP leading to the open conformation of the C-terminus.

Although these results gave insight into what actually happens during the catalytic cycle of the enzyme, they do not provide any conclusive evidence about the mechanism of α3GT. The work thus far did not provide any strong evidence for the previously proposed double displacement mechanism (Gastinel *et al.*, 2001) although subsequent structural and mutational studies on the Glu317Gln mutant did show that residue Glu317 is important in catalysis and the conserved mutation to Gln has reduced both the hydrolase and transferase activity of the enzyme (Zhang *et al.*, 2003). However the magnitude of reduction in activity does not support a double displacement mechanism instead it seems to suggest that Glu317 might play a role in stabilising the cationic transition state for the cleavage of the bond between UDP and C1 of the galactose. This might explain why in the crystal structure of the Glu317Gln mutant in complex with UDP-Gal, a full length UDP-Gal is observed bound in the active site. This complex structure is one of the few crystal structures of glycosyltransferase with an intact active donor substrate bound in the active site. The hydrolysis rate of UDP-Gal by α3GT is very fast and it is notoriously difficult to

obtain a crystal structure of the enzyme in complex with an intact active donor substrate. The mutation to Gln has reduced the hydrolysis and transferase activity to an extent that a complex of the mutant with UDP-Gal managed to be trapped in the crystal structure.

Due to the lack of direct evidence for the double displacement mechanism and the failure to observe any glycosyl-enzyme covalent intermediate in the crystal structures of the active site mutants discussed in section 3.2, it looks that it is highly likely that α3GT might utilise an alternative mechanism. Another detailed work carried out thus far on retaining galactosyltransferases is the characterisation of the structure of *Neisseria meningitides* α-galactosyltransferase LgtC with stable donor and acceptor substrate analogues bound. This study revealed an active site in which the only functional group appropriately positioned to act as a nucleophile was that of the side chain amide of Gln189 (Persson *et al.*, 2001). Replacement of this residue with alanine however, yielded an enzyme that retained sufficient activity (~3%) to shed considerable doubt on its role as a nucleophile (Persson *et al.*, 2001). The authors of this paper went on to suggest that a S_Ni-like mechanism involving a direct attack on the C1 atom of UDP-Gal by the acceptor molecule itself concurrent with cleavage of the glycosidic bond and direct transfer of the Gal moiety to the disaccharide acceptor without the formation of a glycosyl enzyme intermediate. In such a mechanism, the enzyme acts as a scaffold that precisely orients substrates in close proximity, decreasing the energy of the transition state by stabilising the oxocarbenium ion-like species and activating the leaving group (Lairson *et al.*, 2004; Persson *et al.*, 2001). It is postulated that α3GT, a retaining enzyme as well, might carry out a similar catalytic mechanism.

However the evidence to date does not totally support either the double displacement mechanism or the S_Ni-like mechanism and more experimental evidence is clearly needed before a lucid understanding of the mechanism of retaining glycosyltransferase will be obtained. As attempts to obtain a crystal structure of α3GT in complex with both donor and acceptor substrates has yet to be carried out, this could be the next step in trying to understand the catalytic mechanism of the enzyme.

The structure of the wild type α3GT in complex with pNPGal showed the enzyme exhibits some degree of substrate promiscuity. This promiscuity can be useful in utilising α3GT for the synthesis of novel glycoconjugates for therapeutic use. The promiscuity displayed by α3GT is yet unclear and not well characterised. Questions remain on how far this promiscuity can be utilised for synthetic use and this phenomenon coupled with a better catalytic understanding of α3GT needs to be more widely studied in order to harness its synthetic usefulness. There is plenty of scope for further investigation of the use of alternative substrates especially acceptor substrates. The general ease of producing good quality diffracting crystals of α3GT would enable soaking studies with potential acceptor substrates to be carried out. High resolution data is needed in order to clearly observe the mode of binding for these new acceptor molecules and these information can then be subsequently used for the design of novel glycoconjugates or inhibitors.

Crystallisation of Forssman synthase (FS) enzyme is still at an early stage however recent results obtained gave good indications for subsequent crystallisation experiments. The challenge is to produce large diffracting crystals, only showers of needles have thus far resulted from all the conditions that have been identified. More optimisation trials should be carried out by varying the concentration of the protein used, the precipitant concentration and varying the pH of the buffer. Addition of additives should also be carried out and subsequently seeding of the needles into slightly lower precipitant or protein concentrations might be tried to enhance crystal growth.

References

Acharya, K.R., Shapiro, R., Riordan, J.F. and Vallee, B.L. (1995) Crystal structure of bovine angiogenin at 1.5-Å resolution. *Proc. Natl. Acad. Sci. USA*, **92**, 2949-2953.

Basu, M., De, T., Das, K.K., Kyle, J.W., Chon, H.C., Schaeper, R.J. and Basu, S. (1987) Glycolipids. *Methods Enzymol.*, **138**, 575-607.

Bergfors, T.M. (1999) *Protein Crystallization Techniques, Strategies and Tips: A Laboratory Manual*. International University Line, La Jolla, USA.

Bertozzi, C.R. and Kiessling, L.L. (2001) Chemical glycobiology. *Science*, **291**, 2357-2364.

Blow, D. (2002) *Outline of Crystallography for Biologists*. Oxford University Press, Oxford.

Blundell, T.L. and Johnson, L.N. (1976) *Protein Crystallography*. Academic Press Inc. Ltd, London.

Boix, E., Swaminathan, G.J., Z., Y., Natesh, R., Brew, K. and Acharya, K.R. (2002) Structural basis of ordered binding of donor and acceptor substrates to the retaining glycosyltransferase: α-1,3 galactosyltransferase. *J. Biol. Chem.*, **277**, 28310-28318.

Boix, E., Swaminathan, G.J., Zhang, Y., Natesh, R., Brew, K. and Acharya, K.R. (2001) Structure of UDP complex of UDP-Galactose: β-galactoside-α-1,3-galactosyltransferase at 1.53 Å resolution reveals a conformational change in the catalytically important C-terminus. . *J. Biol. Chem.*, **276**, 48608-48614.

Bolam, D.N., Roberts, S., Proctor, M.R., Turkenburg, J.P., Dodson, E.J., Martinez-Fleites, C., Yang, M., Davis, B.G., Davies, G.J. and Gilbert, H.J. (2007) The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proc. Natl .Acad. Sci. USA*, **104**, 5336-5341.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, **72**, 248-254.

Brayer, G.D., Sidhu, G., Maurus, R., Rydberg, E.H., Braun, C., Wang, Y., Nguyen, N.T., Overall, C.M. and Withers, S.G. (2000) Subsite mapping of the human pancreatic alpha-amylase active site through structural, kinetic, and mutagenesis techniques. *Biochemistry*, **39**, 4778-4791.

Breton, C., Bettler, E., Joziasse, D.H., Geremia, R.A. and Imberty, A. (1998) Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases. *J. Biochem. (Tokyo)*, **123**, 1000-1009.

Breton, C. and Imberty, A. (1999) Structure/function studies of glycosyltransferases. *Curr. Opin. Struct. Biol.*, **9**, 563-571.

Breton, C., Snajdrova, L., Jeanneau, C., Koca, J. and Imberty, A. (2006) Structures and mechanisms of glycosyltransferases. *Glycobiology*, **16**, 29R-37R.

Brunger, A.T., Adamas, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kuntsleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. and others. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.*, **D54**, 905-921.

Burkart, M.D., Vincent, S.P., Duffels, A., Murray, B.W., Ley, S.V. and Wong, C.H. (2000) Chemo-enzymatic synthesis of fluorinated sugar nucleotide: useful

mechanistic probes for glycosyltransferases. *Bioorg. Med. Chem.*, **8**, 1937-1946.

Buschiazzo, A., Ugalde, J.E., Guerin, M.E., Shepard, W., Ugalde, R.A. and Alzari, P.M. (2004) Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. *Embo J.*, **23**, 3196-3205.

Carter, P. and Wells, J.A. (1987) Engineering enzyme specificity by "substrate-assisted catalysis". *Science*, **237**, 394-399.

CCP4. (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760-763.

Charnock, S.J. and Davies, G.J. (1999a) Cloning, crystallization and preliminary X-ray analysis of a nucleotide-diphospho-sugar transferase spsA from Bacillus subtilis. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 677-678.

Charnock, S.J. and Davies, G.J. (1999b) Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms. *Biochemistry*, **38**, 6380-6385.

Chen, Z.C., Tanemura, M. and Galili, U. (2001) Synthesis of alpha-gal epitopes (Gal α-3Galβ1-4GlcNAc-R) on human tumor cells by recombinant alpha1,3galactosyltransferase produced in Pichia pastoris. *Glycobiology*, **11**, 577-586.

Chiu, C.P., Watts, A.G., Lairson, L.L., Gilbert, M., Lim, D., Wakarchuk, W.W., Withers, S.G. and Strynadka, N.C. (2004) Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog. *Nat. Struct. Mol. Biol.*, **11**, 163-170.

Compain, P. and Martin, O.R. (2001) Carbohydrate mimetics-based glycosyltransferase inhibitors. *Bioorg. Med. Chem.*, **9**, 3077-3092.

Coutinho, P.M., Deleury, E., Davies, G.J. and Henrissat, B. (2003a) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.*, **328**, 307-317.

Coutinho, P.M., Stam, M., Blanc, E. and Henrissat, B. (2003b) Why are there so many carbohydrate-active enzyme-related genes in plants? *Trends Plant Sci.*, **8**, 563-565.

Daniel, R.M., Dunn, R.V., Finney, J.L. and Smith, J.C. (2003) The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 69-92.

Dauter, Z. (1997) Data collection strategy. *Methods Enzymol.*, **276**, 326-344.

Deriy, L., Ogawa, H., Gao, G.P. and Galili, U. (2005) In vivo targeting of vaccinating tumor cells to antigen-presenting cells by a gene therapy method with adenovirus containing the α-1,3galactosyltransferase gene. *Cancer Gene Ther.*, **12**, 528-539.

Dwek, R.A. (1996) Glycobiology: Toward Understanding the Function of Sugars. *Chem. Rev.*, **96**, 683-720.

Elliott, S.P., Yu, M., Xu, H. and Haslam, D.B. (2003) Forssman synthetase expression results in diminished shiga toxin susceptibility: a role for glycolipids in determining host-microbe interactions. *Infect. Immun.*, **71**, 6543-6552.

Emsley, P., and Cowtan, K. (2004) Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D - Biological Crystallography*, **60**, 2126-2132.

Flint, J., Taylor, E., Yang, M., Bolam, D.N., Tailford, L.E., Martinez-Fleites, C., Dodson, E.J., Davis, B.G., Gilbert, H.J. and Davies, G.J. (2005) Structural dissection and high-throughput screening of mannosylglycerate synthase. *Nat. Struct. Mol. Biol.*, **12**, 608-614.

Franco, O.L. and Rigden, D.J. (2003) Fold recognition analysis of glycosyltransferase families: further members of structural superfamilies. *Glycobiology*, **13**, 707-712.

Fritz, T.A., Hurley, J.H., Trinh, L.B., Shiloach, J. and Tabak, L.A. (2004) The beginnings of mucin biosynthesis: the crystal structure of UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-T1. *Proc. Natl. Acad. Sci. U S A*, **101**, 15307-15312.

Galili, U. (1999) Evolution of α 1,3galactosyltransferase and of the alpha-Gal epitope. *Subcell Biochem.*, **32**, 1-23.

Galili, U. (2005) The α-gal epitope and the anti-Gal antibody in xenotransplantation and in cancer immunotherapy. *Immunol. Cell Biol.*, **83**, 674-686.

Galili, U., Clark, M.R., Shohet, S.B., Buehler, J. and Macher, B.A. (1987) Evolutionary relationship between the natural anti-Gal antibody and the Gal α 1,3Gal epitope in primates. *Proc. Natl. Acad. Sci U S A*, **84**, 1369-1373.

Galili, U., Shohet, S.B., Kobrin, E., Stults, C.L. and Macher, B.A. (1988) Man, apes, and Old World monkeys differ from other mammals in the expression of alpha-galactosyl epitopes on nucleated cells. *J. Biol. Chem.*, **263**, 17755-17762.

Garman, E.F.a.S., T.R. (1997) Macromolecular crystallography. *Journal of Applied Crystallography*, **30**, 211-237.

Gastinel, L.N., Bignon, C., Misra, A.K., Hindsgaul, O., Shaper, J.H. and Joziasse, D.H. (2001) Bovine α1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases. *EMBO J.*, 638-649.

Gastinel, L.N., Cambillau, C. and Bourne, Y. (1999) Crystal structures of the bovine β4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *Embo J.*, **18**, 3546-3557.

Giacovazzo, C., Monaco, H.L., Artioli, G., Viterbo, D., Ferraris, G., Gilli, G., Zanotti, G. and Catti, M. (2002) *Fundamentals of Crystallography*. Oxford University Press, Oxford.

Gibbons, B.J., Roach, P.J. and Hurley, T.D. (2002) Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *J. Mol. Biol.*, **319**, 463-477.

Gibson, R.P., Turkenburg, J.P., Charnock, S.J., Lloyd, R. and Davies, G.J. (2002) Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chem. Biol.*, **9**, 1337-1346.

Goldsmith, E.J., Sprang, S.R., Hamlin, R., Xuong, N.H. and Fletterick, R.J. (1989) Domain separation in the activation of glycogen phosphorylase a. *Science*, **245**, 528-532.

Goodsell, D.S. and Olson, A.J. (1990) Automated docking of substrates to proteins by simulated annealing. *Proteins*, **8**, 195-202.

Gordon, R.D., Sivarajah, P., Satkunarajah, M., Ma, D., Tarling, C.A., Vizitiu, D., Withers, S.G. and Rini, J.M. (2006) X-ray crystal structures of rabbit N-acetylglucosaminyltransferase I (GnT I) in complex with donor substrate analogues. *J. Mol. Biol.*, **360**, 67-79.

Guerin, M.E., Buschiazzo, A., Kordulakova, J., Jackson, M., Alzari, P.M.(2007) Crystal structure of PimA, an essential mannosyltransferase from Mycobacterium smegmatis. *To be Published.*

Gunasekaran, K. and Nussinov, R. (2004) Modulating functional loop movements: the role of highly conserved residues in the correlated loop motions. *Chem. biochem.*, **5**, 224-230.

Ha, S., Walker, D., Shi, Y. and Walker, S. (2000) The 1.9 Å crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci.*, **9**, 1045-1052.

Hakomori, S. (1986) Glycosphingolipids. *Sci. Am.*, **254**, 44-53.

Hammes-Schiffer, S. and Benkovic, S.J. (2006) Relating protein motion to catalysis. *Annu. Rev. Biochem.*, **75**, 519-541.

Hansson, G.C., Karlsson, K.A., Larson, G., Stromberg, N. and Thurin, J. (1985) Carbohydrate-specific adhesion of bacteria to thin-layer chromatograms: a rationalized approach to the study of host cell glycolipid receptors. *Anal. Biochem.*, **146**, 158-163.

Hansson, G.C., Karlsson, K.A., Larson, G., Stromberg, N., Thurin, J., Orvell, C. and Norrby, E. (1984) A novel approach to the study of glycolipid receptors for viruses. Binding of Sendai virus to thin-layer chromatograms. *FEBS Lett.*, **170**, 15-18.

Heissigerova, H., Breton, C., Moravcova, J. and Imberty, A. (2003) Molecular modeling of glycosyltransferases involved in the biosynthesis of blood group A, blood group B, Forssman, and iGb3 antigens and their interaction with substrates. *Glycobiology*, **13**, 377-386.

Helliwell, J.R. (1992) *Macromolecular Crystallography with Synchrotron Radiation.* Cambridge University Press, Oxford.

Helliwell, J.R. (1997) Overview of synchrotron radiation and macromolecular crystallography. *Methods Enzymol.*, **115**, 41-55.

Hendrickson, W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51-58.

Henion, T.R., Macher, B.A., Anaraki, F. and Galili, U. (1994) Defining the minimal size of catalytically active primate α 1,3 galactosyltransferase: structure-function studies on the recombinant truncated enzyme. *Glycobiology*, **4**, 193-201.

Henry, N.F.M. and Lonsdale, K. (1952) *International tables for X-ray Crystallography.* The Kynoch Press.

Horcajada, C., Guinovart, J.J., Fita, I. and Ferrer, J.C. (2006) Crystal structure of an archaeal glycogen synthase: insights into oligomerization and substrate binding of eukaryotic glycogen synthases. *J. Biol. Chem.*, **281**, 2923-2931.

Hu, Y., Chen, L., Ha, S., Gross, B., Falcone, B., Walker, D., Mokhtarzadeh, M. and Walker, S. (2003) Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. *Proc. Natl. Acad. Sci. U S A*, **100**, 845-849.

Jamaluddin, H., Tumbale, P., Withers, S.G., Acharya, K.R. and Brew, K. (2007) Conformational Changes Induced by Binding UDP-2F-galactose to α-1,3 Galactosyltransferase- Implications for Catalysis. *J. Mol. Biol.*, **369**, 1270-1281.

Jank, T., Reinert, D.J., Giesemann, T., Schulz, G.E. and Aktories, K. (2005) Change of the donor substrate specificity of *Clostridium difficile* toxin B by site-directed mutagenesis. *J. Biol. Chem.*, **280**, 37833-37838.

Jinek, M., Chen, Y.W., Clausen, H., Cohen, S.M. and Conti, E. (2006) Structural insights into the Notch-modifying glycosyltransferase Fringe. *Nat. Struct. Mol. Biol.*, **13**, 945-946.

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building models in electron density maps and the location of errors in these models. *Acta cryst. A*, **47**, 110-119.

Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

Kakuda, S., Shiba, T., Ishiguro, M., Tagawa, H., Oka, S., Kajihara, Y., Kawasaki, T., Wakatsuki, S. and Kato, R. (2004) Structural basis for acceptor substrate recognition of a human glucuronyltransferase, GlcAT-P, an enzyme critical in the biosynthesis of the carbohydrate epitope HNK-1. *J. Biol. Chem.*, **279**, 22693-22703.

Karlsson, K.A. and Stromberg, N. (1987) Overlay and solid-phase analysis of glycolipid receptors for bacteria and viruses. *Methods Enzymol.*, **138**, 220-232.

Klein, H.W., Im, M.J. and Palm, D. (1986) Mechanism of the phosphorylase reaction. Utilization of D-gluco-hept-1-enitol in the absence of primer. *Eur. J. Biochem.*, **157**, 107-114.

Kleywegt, G.J. and Jones, T.A. (1998) Databases in protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 1119-1131.

Kobayashi, T. and Cooper, D.K. (1999) Anti-Gal, alpha-Gal epitopes, and xenotransplantation. *Subcell Biochem.*, **32**, 229-257.

Kubo, A., Arai, Y., Nagashima, S. and Yoshikawa, T. (2004) Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. *Arch. Biochem. Biophys.*, **429**, 198-203.

Kuwaki, K., Tseng, Y.L., Dor, F.J., Shimizu, A., Houser, S.L., Sanderson, T.M., Lancos, C.J., Prabharasuth, D.D., Cheng, J., Moran, K., Hisashi, Y., Mueller, N., Yamada, K., Greenstein, J.L., Hawley, R.J., Patience, C., Awwad, M., Fishman, J.A., Robson, S.C., Schuurman, H.J., Sachs, D.H. and Cooper, D.K. (2005) Heart transplantation in baboons using α1,3-galactosyltransferase gene-knockout pigs as donors: initial experience. *Nat. Med.*, **11**, 29-31.

Lairson, L.L., Chiu, C.P., Ly, H.D., He, S., Wakarchuk, W.W., Strynadka, N.C. and Withers, S.G. (2004a) Intermediate trapping on a mutant retaining α-galactosyltransferase identifies an unexpected aspartate residue. *J. Biol. Chem.*, **279**, 28339-28344.

Lairson, L.L. and Withers, S.G. (2004b) Mechanistic analogies amongst carbohydrate modifying enzymes. *Chem. Commun. (Camb)*, 2243-2248.

Lairson, L.L., Watts, A.G., Wakarchuk, W.W. and Withers, S.G. (2006) Using substrate engineering to harness enzymatic promiscuity and expand biological catalysis. *Nat. Chem. Biol.*, **2**, 724-728.

Lanteri, M., Giordanengo, V., Vidal, F., Gaudray, P. and Lefebvre, J.C. (2002) A complete α1,3-galactosyltransferase gene is present in the human genome and partially transcribed. *Glycobiology*, **12**, 785-792.

Lariviere, L., Sommer, N. and Morera, S. (2005) Structural evidence of a passive base-flipping mechanism for AGT, an unusual GT-B glycosyltransferase. *J. Mol. Biol.*, **352**, 139-150.

Laskowski, R.A., Moss, D.S. and Thornton, J.M. (1993) Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.*, **231**, 1049-1067.

Lazarus, B.D., Milland, J., Ramsland, P.A., Mouhtouris, E. and Sandrin, M.S. (2002) Histidine 271 has a functional role in pig α-1,3galactosyltransferase enzyme activity. *Glycobiology*, **12**, 793-802.

Lesk, A.M. (1995) NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.*, **5**, 775-783.

Leslie, A.G.W. (2001) *Integration of macromolecular diffraction data. Intenational tables for crystallography. M.G. Rossman and E. Arnold.* Kluwer Academic Publishers.

Liu, J. and Mushegian, A. (2003) Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.*, **12**, 1418-1431.

Lobsanov, Y.D., Romero, P.A., Sleno, B., Yu, B., Yip, P., Herscovics, A. and Howell, P.L. (2004) Structure of Kre2p/Mnt1p: a yeast α1,2-mannosyltransferase involved in mannoprotein biosynthesis. *J. Biol. Chem.*, **279**, 17921-17931.

Lovering, A.L., De Castro, L., Lim, D. and Strynadka, N.C. (2006) Structural analysis of an "open" form of PBP1B from *Streptococcus pneumoniae*. *Protein Sci.*, **15**, 1701-1709.

Mackiewicz, A., Dewey, M.J., Berger, F.G. and Baumann, H. (1991) Acute phase mediated change in glycosylation of rat α 1-acid glycoprotein in transgenic mice. *Glycobiology*, **1**, 265-269.

Manches, O., Plumas, J., Lui, G., Chaperot, L., Molens, J.P., Sotto, J.J., Bensa, J.C. and Galili, U. (2005) Anti-Gal-mediated targeting of human B lymphoma cells to antigen-presenting cells: a potential method for immunotherapy using autologous tumor cells. *Haematologica*, **90**, 625-634.

Mangani, S. and Liljas, A. (1993) Crystal structure of the complex between human carbonic anhydrase II and the aromatic inhibitor 1,2,4-triazole. *J. Mol. Biol.*, **232**, 9-14.

Martinez-Fleites, C., Proctor, M., Roberts, S., Bolam, D.N., Gilbert, H.J. and Davies, G.J. (2006) Insights into the synthesis of lipopolysaccharide and antibiotics

through the structures of two retaining glycosyltransferases from family GT4. *Chem. Biol.*, **13**, 1143-1152.

Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 479-491.

McDonald, I.K. and Thornton, J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777-793.

Monegal, A. and Planas, A. (2006) Chemical Rescue of α3-Galactosyltransferase. Implications in the Mechanism of Retaining Glycosyltransferases. *J. Am. Chem. Soc.*, **128**, 16030-16031.

Monner, D.A. and Muhlradt, P.F. (1993) Surface expression of Forssman glycosphingolipid antigen on murine bone marrow-derived macrophages is subject to both temporal and population-specific regulation and is modulated by IL-4 and IL-6. *Immunobiology*, **188**, 82-98.

Mulichak, A.M., Losey, H.C., Lu, W., Wawrzak, Z., Walsh, C.T. and Garavito, R.M. (2003) Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc. Natl. Acad. Sci. U S A*, **100**, 9238-9243.

Mulichak, A.M., Losey, H.C., Walsh, C.T. and Garavito, R.M. (2001) Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure*, **9**, 547-557.

Mulichak, A.M., Lu, W., Losey, H.C., Walsh, C.T. and Garavito, R.M. (2004) Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. *Biochemistry*, **43**, 5170-5180.

Murray, B.W., Takayama, S., Schultz, J. and Wong, C.H. (1996) Mechanism and specificity of human α-1,3-fucosyltransferase V. *Biochemistry*, **35**, 11183-11195.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240-255.

Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157-163.

Nevskaya, N., Tischenko, S., Fedorov, R., Al-Karadaghi, S., Liljas, A., Kraft, A., Piendl, W., Garber, M. and Nikonov, S. (2000) Archaeal ribosomal protein L1: the structure provides new insights into RNA binding of the L1 protein family. *Structure*, **8**, 363-371.

Ni, L., Sun, M., Yu, H., Chokhawala, H., Chen, X. and Fisher, A.J. (2006) Cytidine 5'-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from Pasteurella multocida. *Biochemistry*, **45**, 2139-2148.

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B.G., Tarling, C.A., Ford, C.M., Bowles, D.J. and Davies, G.J. (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *Embo J.*, **25**, 1396-1405.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology*, **276**, 307-326.

Pak, J.E., Arnoux, P., Zhou, S., Sivarajah, P., Satkunarajah, M., Xing, X. and Rini, J.M. (2006) X-ray crystal structure of leukocyte type core 2 beta1,6-N-acetylglucosaminyltransferase. Evidence for a convergence of metal ion-independent glycosyltransferase mechanism. *J. Biol. Chem.*, **281**, 26693-26701.

Patenaude, S.I., Seto, N.O., Borisova, S.N., Szpacenko, A., Marcus, S.L., Palcic, M.M. and Evans, S.V. (2002) The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nat. Struct. Biol.*, **9**, 685-690.

Pedersen, L.C., Dong, J., Taniguchi, F., Kitagawa, H., Krahn, J.M., Pedersen, L.G., Sugahara, K. and Negishi, M. (2003) Crystal structure of an α 1,4-N-acetylhexosaminyltransferase (EXTL2), a member of the exostosin gene family involved in heparan sulfate biosynthesis. *J. Biol. Chem.*, **278**, 14420-14428.

Pedersen, L.C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T.A. and Negishi, M. (2000) Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I. *J. Biol. Chem.*, **275**, 34580-34585.

Persson, K., Ly, H.D., Dieckelmann, M., Wakarchuk, W.W., Withers, S.G. and Strynadka, N.C. (2001) Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat. Struct. Biol.*, **8**, 166-175.

Persson, M., Letts, J.A., Hosseini-Maaf, B., Borisova, S.N., Palcic, M.M., Evans, S.V. and Olsson, M.L. (2007) Structural effects of naturally occurring human blood group B galactosyltransferase mutations adjacent to the DXD motif. *J. Biol. Chem.*, **282**, 9564-9570.

Perutz, M.F. (1956) Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Cryst.* , **9**, 867-873.

Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, **7**, 95-99.

Ramakrishnan, B., Boeggeman, E. and Qasba, P.K. (2005) Mutation of arginine 228 to lysine enhances the glucosyltransferase activity of bovine β-1,4-galactosyltransferase I. *Biochemistry*, **44**, 3202-3210.

Ramakrishnan, B., Boeggeman, E., Ramasamy, V. and Qasba, P.K. (2004) Structure and catalytic cycle of β-1,4-galactosyltransferase. *Curr. Opin. Struct. Biol.*, **14**, 593-600.

Ramakrishnan, B. and Qasba, P.K. (2001) Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the β1,4-galactosyltransferase-I. *J. Mol. Biol.*, **310**, 205-218.

Ramakrishnan, B. and Qasba, P.K. (2002) Structure-based design of beta 1,4-galactosyltransferase I (beta 4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity: point mutation broadens wβ 4Gal-T1 donor specificity. *J. Biol. Chem.*, **277**, 20833-20839.

Ramakrishnan, B., Ramasamy, V. and Qasba, P.K. (2006) Structural snapshots of beta-1,4-galactosyltransferase-I along the kinetic pathway. *J. Mol. Biol.*, **357**, 1619-1633.

Read, R.J. (2001) Pushing boundaries of molecular replacement with maximum likelihood. *Acta Cryst. D*, **57**, 1373-1382.

Reinert, D.J., Jank, T., Aktories, K. and Schulz, G.E. (2005) Structural basis for the function of *Clostridium difficile* toxin B. *J. Mol. Biol.*, **351**, 973-981.

Rhodes, G. (2000) *Crystallography Made Crystal Clear*. Elsevier Science (USA).

Ritchie, D.W. and Kemp, G.J.L. (2000) Protein Docking Using Spherical Polar Fourier Correlations. *Proteins: Struct. Func. Genet.*, **39**, 178-194.

Rodgers, D.W. (1994) Crystallography. *Structure*, **2**, 1135-1140.

Rossman, M.G. (1972) *The molecular replacement method*. Gordon and Breach, New York.

Rudd, P.M., Elliott, T., Cresswell, P., Wilson, I.A. and Dwek, R.A. (2001) Glycosylation and the immune system. *Science*, **291**, 2370-2376.

Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C. (1997) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.*, **23**, 327-341.

Sadahira, Y., Yasuda, T. and Kimoto, T. (1991) Regulation of Forssman antigen expression during maturation of mouse stromal macrophages in haematopoietic foci. *Immunology*, **73**, 498-504.

Sandhoff, K. and van Echten, G. (1993) Ganglioside metabolism--topology and regulation. *Adv. Lipid Res.*, **26**, 119-142.

Shao, H., He, X., Achnine, L., Blount, J.W., Dixon, R.A. and Wang, X. (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell*, **17**, 3141-3154.

Sheldrick, G.M., Dauter, Z., Wilson, K.S., Hope, H. and Sieker, L.C. (1993) The application of direct methods and Patterson interpretation to high-resolution native protein data. *Acta Crystallogr. D Biol. Crystallogr.*, **49**, 18-23.

Simon, P.M., Neethling, F.A., Taniguchi, S., Goode, P.L., Zopf, D., Hancock, W.W. and Cooper, D.K. (1998) Intravenous infusion of Galα1-3Gal oligosaccharides in baboons delays hyperacute rejection of porcine heart xenografts. *Transplantation*, **65**, 346-353.

Sippl, M.J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins*, **17**, 355-362.

Snajdrova, L., Kulhanek, P., Imberty, A. and Koca, J. (2004) Molecular dynamics simulations of glycosyltransferase LgtC. *Carbohydr. Res.*, **339**, 995-1006.

Springer, T.A. (1990) Adhesion receptors of the immune system. *Nature*, **346**, 425-434.

Storoni, L.C., Mccoy, A.J. and Read, R.J. (2004) Likelihood-enhanced fast rotation functions. *Acta cryst. D*, **60**, 432-438.

Stults, C.L., Sweeley, C.C. and Macher, B.A. (1989) Glycosphingolipids: structure, biological source, and properties. *Methods Enzymol.*, **179**, 167-214.

Svanborg Eden, C., Andersson, B., Hagberg, L., Hanson, L.A., Leffler, H., Magnusson, G., Noori, G., Dahmen, J. and Soderstrom, T. (1983) Receptor analogues and anti-pili antibodies as inhibitors of bacterial attachment in vivo and in vitro. *Ann. N Y Acad. Sci.*, **409**, 580-592.

Takayama, S., Chung, S.J., Igarashi, Y., Ichikawa, Y., Sepp, A., Lechler, R.I., Wu, J., Hayashi, T., Siuzdak, G. and Wong, C.H. (1999) Selective inhibition of β-

1,4- and α-1,3-galactosyltransferases: donor sugar-nucleotide based approach. *Bioorg. Med. Chem.*, **7**, 401-409.

Tarbouriech, N., Charnock, S.J. and Davies, G.J. (2001) Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDP-sugar glycosyltransferases. *J. Mol. Biol.*, **314**, 655-661.

Teodoro, M.L., Phillips, G.N., Jr. and Kavraki, L.E. (2003) Understanding protein flexibility through dimensionality reduction. *J. Comput. Biol.*, **10**, 617-634.

Thoden, J.B. and Holden, H.M. (1998) Dramatic differences in the binding of UDP-galactose and UDP-glucose to UDP-galactose 4-epimerase from *Escherichia coli*. *Biochemistry*, **37**, 11469-11477.

Unligil, U.M. and Rini, J.M. (2000) Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.*, **10**, 510-517.

Unligil, U.M., Zhou, S., Yuwaraj, S., Sarkar, M., Schachter, H. and Rini, J.M. (2000) X-ray crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. *Embo J.*, **19**, 5269-5280.

Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.*, **30**, 1022-1025.

Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, **3**, 97-130.

Vrielink, A., Ruger, W., Driessen, H.P. and Freemont, P.S. (1994) Crystal structure of the DNA modifying enzyme β-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *Embo J.*, **13**, 3413-3422.

Wells, L., Vosseller, K. and Hart, G.W. (2001) Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science*, **291**, 2376-2378.

Xu, H., Storch, T., Yu, M., Elliott, S.P. and Haslam, D.B. (1999) Characterization of the human Forssman synthetase gene. An evolving association between glycolipid synthesis and host-microbial interactions. *J. Biol. Chem.*, **274**, 29390-29398.

Yamamoto, F., Clausen, H., White, T., Marken, J. and Hakomori, S. (1990) Molecular genetic basis of the histo-blood group ABO system. *Nature*, **345**, 229-233.

Yamamoto, F. and McNeill, P.D. (1996) Amino acid residue at codon 268 determines both activity and nucleotide-sugar donor substrate specificity of human histo-blood group A and B transferases. In vitro mutagenesis study. *J. Biol. Chem.*, **271**, 10515-10520.

Yamamoto, M., Lin, X.H., Kominato, Y., Hata, Y., Noda, R., Saitou, N. and Yamamoto, F. (2001) Murine equivalent of the human histo-blood group ABO gene is a cis-AB gene and encodes a glycosyltransferase with both A and B transferase activity. *J. Biol. Chem.*, **276**, 13701-13708.

Yazer, M.H. and Palcic, M.M. (2005) The importance of disordered loops in ABO glycosyltransferases. *Transfus. Med. Rev.*, **19**, 210-216.

Yuan, Y., Barrett, D., Zhang, Y., Kahne, D., Sliz, P. and Walker, S. (2007) Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. *Proc. Natl. Acad. Sci. USA*, **104**, 5348-5353.

Zhang, Y., Deshpande, A., Xie, Z., Natesh, R., Acharya, K.R. and Brew, K. (2004) Roles of active site tryptophans in substrate binding and catalysis by α-1,3 galactosyltransferase. *Glycobiology*, **14**, 1295-1302.

Zhang, Y., Swaminathan, G.J., Deshpande, A., Natesh, R., Boix, E., Xie, Z., Acharya, K.R. and Brew, K. (2003) Roles of individual enzyme-substrate interactions by α-1,3 galactosyltransferase in catalysis and specificity. *Biochemistry*, 13512-13521.

Zhang, Y., Wang, P.G. and Brew, K. (2001) Specificity and mechanism of metal ion activation in UDP-galactose β-galactoside α 1,3-galactosyltransferase. *J. Biol. Chem*, **276**, 11567-11574.